Profiling for Success: Reasoning Tests

User's Guide v1.3

Roy Childs, John Gosling, Mark Parkinson, Angus S. McDonald

Contents

Introduction 1

Section One: Using Reasoning Tests for selection and development 3

- Why use Reasoning Tests? 3
- Reasons for developing the PfS-Reasoning Tests 7

Section Two: Selecting and administering the Reasoning Tests 11

- Introduction 11
- Selecting appropriate tests 11
- Using the PfS- Reasoning Tests as locator tests 13
- Administering paper-based tests 15
- Overview of administration 15
- Planning the test session 16
- Materials 17
- The test session 18
- Administering computer-based tests 19
- Supervised assessment 20
- Unsupervised assessment 21
- Technical requirements for computer-based tests 23

Section Three: Scoring and review of test results 25

- Overview of scoring and test scores 25
- Qualitative analysis of results 26
- Scoring paper-based tests 27
- Scoring computer-based tests 29
- Using the online report generator with paper-based tests 29
- Review of test results 29
- Communicating test results 30
- Conducting a review session 32

Section Four: Development of the Reasoning Tests 35

- Test formats 35
- Verbal reasoning format 36
- Numerical reasoning format 36
- Abstract reasoning format 37
- Item writing 38
- Pre-trialling item reviews 38
- Trialling 39
- Item analysis 39

© 2003 – 2013 Team Focus Limited

Section Five: Technical information 45

Introduction 45
Reliability 45
The concept of reliability 45
Reliability statistics 46
Standard error of difference 51
Bias 53
A commentary on interpreting bias data 61
Validity 62
Face validity 62
Content validity 63
Construct validity 63
Criterion validity 69
References 73

Appendix One: Explanations of practice questions 75

Appendix Two: Sample test reports 81

Appendix Three: Norm tables 89

Introduction to the norm tables 89
General norms for closed tests 90
Descriptions of norms for open tests 103
Additional norms for closed tests - Supplement 1 107
Descriptions of additional norms for open tests - Supplement 1 119
Additional norms for closed tests - Supplement 2 121

Appendix 4: Comparison tables 151

© 2003 – 2013 Team Focus Limited

List of tables

Table 1: Correspondence between the PfS-Reasoning Tests and level of ability as indicated by the level of educational attainment

Table 2: Appropriate Verbal, Numerical and Abstract test levels for locator test percentile scores

Table 3: Summary descriptions for combinations of speed of working and accuracy

Table 4: Score bands used in summary reports and their relationship to T-scores and percentiles

Table 5: Timings and number of items in each of the PfS- Reasoning Tests

Table 6: Mean, SD, sample size, number of items, internal consistency and SEM for the PfS-Reasoning Tests

Table 7: Mean and SD for first time and retest candidates, and test-retest reliabilities for bespoke versions of the PfS-Reasoning Tests

Table 8: Difficulty levels for the closed PfS-Reasoning Tests and parallel form reliability

Table 9: Mean raw scores and standard deviations for males and females on the PfS-Reasoning Tests

Table 10: Mean raw scores and standard deviations for ‘whites’ and ‘non-whites’ on the PfS-Reasoning Tests

Table 11: Mean test scores and effect sizes for different ethnic groups based on the open Level 2 PfS-Reasoning Tests Reasoning Tests

Table 12: Cross tabulation between appointment decision and aggregated ethnic group differences

Table 13: Raw score means on the three reasoning tests for each ethnic group

Table 14: Associations between raw PfS-Reasoning Tests and respondents age

Table 15: Intercorrelations of the PfS-Reasoning Tests

Table 16: Associations between PfS-Reasoning Tests and the GMAT

Table 17: Associations between PfS Abstract Tests and GMA Abstract form A

Table 18: Inter-correlations between the Verbal, Numerical and Abstract Reasoning Tests and existing reasoning tests

© 2003 – 2013 Team Focus Limited

Table 19: Correlations between Verbal, Numerical and Abstract Reasoning Tests and Memory and Attention Test

Table 20: Associations between GCSE English, maths and science grades and PfS-Reasoning Tests

Table 21: The association between UCAS points, degree class and PfS-Reasoning Tests

List of figures

Figure 1: The predictive validity and popularity of different assessment methods

Figure 3: PfS-Reasoning Test levels and summary information

Figure 3: The normal distribution curve, Z-score, T-score and percentile scales

Figure 4: Test characteristic curves for Verbal Tests 1-4

Figure 5: Test characteristic curves for Numerical Tests 1-4

Figure 6: Test characteristic curves for Abstract Tests 1-4

© 2003 – 2013 Team Focus Limited

Introduction

The Profiling for Success-Reasoning Tests (PfS-Reasoning Tests) offer a flexible approach to the assessment of reasoning abilities for selection and development purposes. The tests cover three areas of reasoning abilities:

Verbal – The ability to understand written information and determine what follows logically from the information.

Numerical – The ability to use numerical information to solve problems.

Abstract – The ability to identify patterns in abstract shapes and generate and test hypotheses.

As the benefits of psychometric assessments are increasingly recognised and test usage grows, new ways of assessing abilities are needed. The PfS-Reasoning Tests meet these needs by offering both paper- and computer-based assessments that can be used with a wide range of ability groups.

The key features and benefits of the PfS-Reasoning Tests are:

- Flexible delivery options – the paper- and computer-based (online) tests allow for traditional individual or group administration, or remote assessment. Through remote assessment it is possible to include test data earlier on in the assessment process. For example, including test information in the first sift alongside application forms or CVs gives more information on which to base decisions, so potentially enhancing the accuracy of decisions and increasing the efficiency of the selection process.

- 'Open' and 'closed' test versions – closed versions of each of the tests are available for use under supervised conditions where the identity of tests takers can be closely monitored. Open access versions are also available for use in situations where remote, unsupervised administration is appropriate. These different versions have been developed to meet the increasing need to test candidates remotely, largely as a result in the growth of internet assessment, and the demand for the use of tests for guidance and other development purposes, as well as the more established approach of supervised assessment.

- Common formats across a wide ability range – the Verbal, Numerical and Abstract Reasoning Tests span a wide range of ability levels, from school leavers to experienced managers, using common test formats. If necessary, the appropriate test level can be identified by administering a test as a 'locator' among a group of current employees. This process is readily achieved through the use of online tests and guidance is given on how to do this in the User's Guide.

© 2003 – 2013 Team Focus Limited

- Confidence that the different levels of the tests are measuring different levels of ability. This has been established by conducting a study using Item Response Theory (IRT) methodology. This reinforces the value of using different levels of tests for different purposes, and also provides a way of equating test scores across different levels of test.

- Detailed reports and analysis – separate computer-generated reports are available for test users and test takers. For test takers these reports give raw and standardised test scores and an analysis of speed and accuracy, linked to a narrative suggesting areas for consideration and development. Test users' reports present full test data and an analysis of speed and accuracy linked to interview prompts. Summary versions of reports for test takers and test users are also available.

This User's Guide provides test users with the information they need to understand, use and interpret the Verbal, Numerical and Abstract Reasoning Tests which make up the PfS-Reasoning Tests. **Section One** summarises research on the importance of reasoning abilities for successful job performance and training, and describes the rationale behind the PfS-Reasoning Tests. Administration of paper- and computer-based versions of the tests is covered in **Section Two**. **Section Three** deals with scoring and feedback. The development of the PfS-Reasoning Tests is described in **Section Four**, and **Section Five** provides technical information on the tests and their functioning. It is recommended that all users should read at least **Sections Two** and **Three** before using any of the tests.

In addition to the information contained in this User's Guide, the test publishers offer consultancy, training and general support in using and interpreting the results from these Reasoning Tests and other assessments. For enquiries and support, please contact Team Focus Ltd on + 44 (0)1628 637338, e-mail teamfocus@teamfocus.co.uk.

© 2003 – 2013 Team Focus Limited

Section One: Using Reasoning Tests for selection and development

Why use Reasoning Tests?

The use of reasoning tests for selection and development is well-established in many organisations. Surveys show that usage continues to increase (e.g. CIPD, 2006; Jenkins, 2001), with new organisations discovering the benefits that properly applied psychometrics can bring and established users expanding their use of psychometrics. The use of online tests as part of the selection process has also grown rapidly in recent years, with figures showing a rise from 6% in 2002 to 25% in 2006 (CIPD, 2004; 2006).

When used sensitively, with due regard for both their strengths and limitations, there are many good reasons for using psychometric tests. The most compelling reason for using psychometrics is that they provide accurate information on a person's potential or development needs. All benefits of psychometric assessments ultimately feed into this accuracy, so helping the decision-making or development process. Well-informed decisions, in turn, help organisations to grow and develop. It is now well-established that tests of general mental ability, of which reasoning is a core feature, are an important factor in the decision-making process, as they are the best single predictor of job performance and success on work-related training courses (Schmidt and Hunter, 1998).

To contribute to the decision-making process, psychometric tests have to discriminate among the people who take them. Here, discrimination is the ability to identify real differences between test takers' potential, not the pejorative sense of discrimination where one group is favoured over another for reasons unrelated to true potential.

Changes in the education system, particularly the increasing number of students in further and higher education, have made psychometric tests valuable decision-making tools for employers for three reasons:

- The growth in the number of courses and qualifications makes it difficult to evaluate applicants with very different qualifications.

- The increasing number of students obtaining top grades means that academic qualifications have lost much of their ability to discriminate between people.

- Standards of education vary considerably between institutions and courses. Psychometric tests overcome these variations by providing a 'level playing-field' for people to demonstrate their current ability and potential.

This last point touches on the increasingly important issue of fairness in selection. A very significant reason for using psychometrics is that they can provide a fair assessment of all applicants. To be fair, the abilities assessed by the test must be related to job performance (see page 9) and administration standardised for all test takers (see Section Two).

© 2003 – 2013 Team Focus Limited

Helping test takers to prepare for the testing session, for example by sending out the Test Taker's Guide (see page 14) or giving access to other approved practice materials, also helps to give everyone an equal chance to demonstrate their abilities.

Psychometric tests further contribute to effective selection and development decisions by explicitly recognising the potential for error in test scores. All assessments (e.g. educational qualifications, ratings from assessment centres or interviews) are subject to error, but this error is rarely acknowledged (see pages 36 and 37 for further discussion of test error). Recognising that test scores contain a degree of error and making this explicit, allows the band of error to be taken into account when making decisions based on test scores.

The relationship between test scores and subsequent job performance or success on training courses has been touched on above. To be defensible as a selection method, links between test scores and subsequent job or training performance have to be established. When this link is established, a test or other selection method is said to have 'validity' or to be 'fit for the purpose'. Showing a test has validity is also important as it is the basis for showing a selection process to be defensible from a legal perspective.

Early research on the links between test scores and subsequent job performance produced mixed results, often due to the limitations of the research itself. More rigorous research methods have since identified a considerable relationship between performance on the family of tests represented by the PfS-Reasoning Tests and job performance (e.g. Bertua, Anderson and Salgado, 2005; Schmidt and Hunter, 1998).

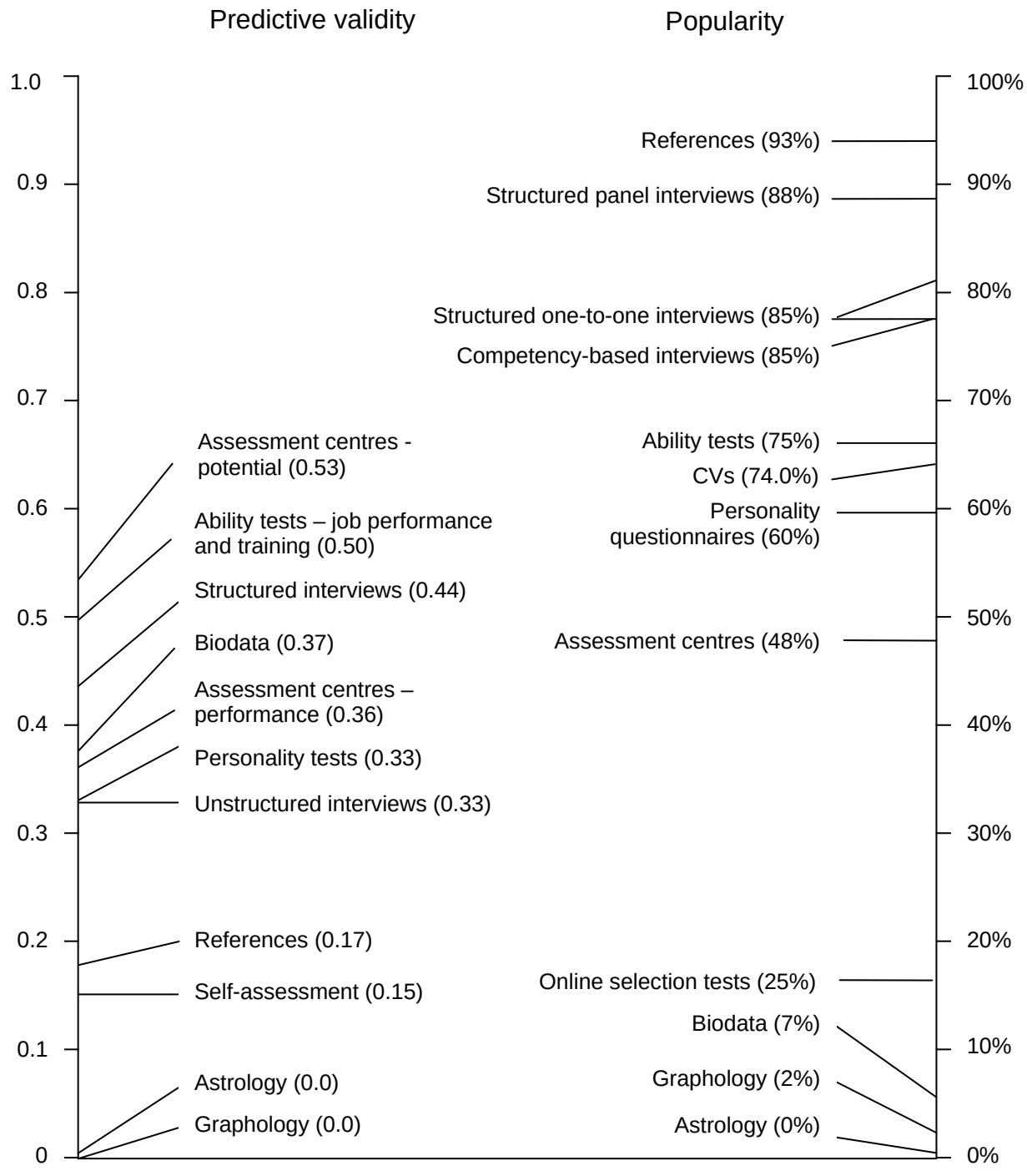
Figure 1 summarises the findings from a number of sources on the predictive validity and popularity of a variety of assessment methods. From this it can be seen that ability tests are very good predictors of job performance and job related training success, and are one of the most frequently used assessment methods after interviews and references.

In a meta-analysis of validity data, Schmidt and Hunter (1988) showed tests of general mental ability to have a predictive validity of 0.51. Recent work using validity studies from the UK produced figures of 0.48 for the relationship between general mental ability and job performance, and 0.50 for the relationship with job related training success (Bertua *et al*, 2005). Although assessment centres have a slightly higher predictive validity, reasoning tests are often administered as part of assessment centres. The incremental validity of assessment centres once ability tests have been allowed for is quite modest, with recent estimates suggesting that, at best, they add no more than 0.1 to the correlation with job performance (Schmidt and Hunter, 1998).

© 2003 – 2013 Team Focus Limited

A further finding of note from Bertua *et al* (2005) was the relationship between predictive validity and different occupational groups. Tests of general mental ability showed higher validities for the 'managerial' and 'professional' categories, indicating their importance for the prediction of more complex, cognitively demanding roles. As the authors note, this finding contradicts the assumption held by some that ability tests have less validity for more senior appointments.

Figure 1: The predictive validity and popularity of different assessment methods



© 2003 – 2013 Team Focus Limited

Notes: figures for predictive validity taken from Beruta, Anderson and Salgado (2005), Gaugler, Rosenthal, Thornton and Bentson (1987), Hunter and Hunter (1982), McDaniel, Whetzel, Schmidt and Maurer (1994), Reilly and Chao (1982), Robertson and Kinder (1993), Schmidt and Hunter (1998). Figures for popularity based on British organisations and taken from CIPD (2000; 2006) and Shackleton and Newell (1991) and indicate use by organisations for at least some appointments.

© 2003 – 2013 Team Focus Limited

Reasons for developing the PfS-Reasoning Tests

The PfS-Reasoning Tests were primarily developed to:

- meet the demand for new materials, due to the increase in psychometric testing;
- offer test users the advantages of computer-based assessments;
- give users the options of 'open' and 'closed' versions of tests assessing the same constructs to address the need for both supervised and unsupervised assessment; and
- deal with common issues with psychometric testing – issues that have been identified from extensive experience of using and training people to use psychometric assessments. These include having sufficient choice to select tests of appropriate level of difficulty according to the ability range of the individuals being tested and to have a consistent format across these levels in order to reduce the learning curve (sometimes needed since tests can have very different formats) for both administrator and testee.

The advantages of computerised testing have been recognised for some time, particularly in the areas of administration, scoring and generating reports (e.g. Kline, 2000). Within the fields of occupational, clinical and educational assessment, computer-based testing is now widely accepted. Test developers are starting to go beyond computerised versions of pencil-and-paper assessments and explore how technology can be used to create innovative and engaging new forms of assessment.

When developing the PfS-Reasoning Tests, one of the goals was to make the potential benefits of testing more accessible. Using the internet as a method of delivery means that the psychometric assessments can be used more flexibly and with a wider range of people. With computer-based testing it is also easier to make tests such as the PfS-Reasoning Tests visually appealing. This is important when assessing people who may have lower levels of motivation for completing the tests, as it makes the testing experience different from traditional educational assessments.

The PfS-Reasoning Tests also meet the need for administrators to be able to set up and monitor the assessment process, to have control of the data and how it is used, and to generate informative reports from the test results. Through the PfS online system, administrators can control which tests are made available to test takers, which norm groups the results are compared to, what types of report are generated and who receives the reports. Security is guaranteed by the use of passwords. Computerised testing makes scoring fast and accurate.

The output from the PfS-Reasoning Tests provides some added value information which goes beyond the usual raw and standardised test scores. They provide an analysis of speed and accuracy (see **Section Three**) to enable the interpreter to consider potential reasons for low or high scores which may have to do with strategy rather than just ability.

© 2003 – 2013 Team Focus Limited

These statistics are combined into reports that give development suggestions and interview prompts, through which reviewers can explore with test takers the meaning and implications of their results. These analysis and reporting facilities mean that all test takers can receive valuable, personalised feedback, regardless of the outcome of the test results. This makes the PfS-Reasoning Tests truly developmental. By using the data entry and scoring facilities of the PfS online assessment system, users of the paper-based tests can also benefit from the features of the automated tests reports (see pages 24 and 25).

Two related challenges often faced by test users are:

- selecting the appropriate ability level of psychometric tests; and
- familiarising themselves with the formats of different tests.

Both of these issues are addressed by the PfS-Reasoning Tests.

It is not usually possible to adequately cover a wide ability range with a single test. Tests provide maximum information when the mean score is close to the middle of the possible score range. A single test capable of assessing accurately across the full ability range would take too long to administer to be practical in organisational settings and would be frustrating for many test takers: able test takers would become bored with many simple questions and less able test takers frustrated at the number of questions they found too difficult. To deal with the issue of ability, there are four levels of the Verbal, Numerical and Abstract Reasoning Tests, spanning school-leavers to people with postgraduate qualifications and considerable professional experience.

Each of the three tests uses a common format. This means that once users have familiarised themselves with one level of a test, they will be equally familiar with all levels of both the closed an open versions. Users of the PfS-Reasoning Tests therefore no longer have to become familiar with different test formats for different populations – simplifying the administration process. The same formats are also used for scoring and reporting, reducing the possibility of errors and making the interpretation and review of test results easier.

Test users often see identifying the appropriate test level as a major challenge, particularly when the test is to be used by a diverse group of people. The PfS-Reasoning Tests address this issue through suggesting how the tests can be used as ‘locator’ tests. By administering one of the tests to an existing group of employees, the results can be used to determine which of the four test levels is appropriate for the position in question. The use and interpretation of locator tests is simplified if the computer-based versions are used. Guidance on how to use and interpret locator tests is given on pages 10 to 12.

© 2003 – 2013 Team Focus Limited

In areas such as graduate and managerial selection and development, the use of psychometrics is well-established. As more organisations use psychometrics there is a risk that the tests become over-exposed, with applicants in some cases taking the same test more than once, so giving them an unfair advantage over others.

All new tests offer a short-term solution to the problem of over-exposure, though this has become an increasingly important issue with the advent of unsupervised testing over the internet. The PfS-Reasoning Tests have also been developed with the goal of addressing this in the long-term. The open and closed versions of the Verbal, Numerical and Abstract Reasoning Tests have been developed to give users confidence in the security of the closed tests whilst retaining the option of unsupervised internet assessment using the open versions. Further parallel versions of the Verbal, Numerical and Abstract Reasoning Tests are already under development and there is also the option for bespoke assessments consisting of unique series of items to be developed for clients on request. The PfS-Reasoning Tests therefore offer organisations the opportunity to avoid the problems associated with the over-exposure of tests.

To summarise the PfS-Reasoning Tests available:

- Levels 1 to 4 closed tests cover the areas of Verbal, Numerical and Abstract Reasoning and are intended to be used for secure testing situations which are either supervised or where they are administered online to known test takers.

- Each level contains a unique set of items and levels are broadly tied to educational stages: Level 1 for test takers in the last years of compulsory education (years 10 and 11), Level 2 for those in further education, Level 3 for undergraduates and Level 4 for postgraduates and experienced professionals.

- Levels 1 and 2 of the open tests are intended for use under less secure conditions (e.g. during an initial sift where results are collected remotely).

- As with the closed tests, each level contains a unique set of items. Level 1 of the open tests covers the same ability range as Levels 1 and 2 of the closed tests and Level 2 of the open tests the same range as Levels 3 and 4 of the closed tests.

- The Combined Reasoning Test consists of items from the Level 1 open tests. As such it is intended for use under less secure conditions, particularly initial sifts and career development or guidance in younger test takers.

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Section Two: Selecting and administering the Reasoning Tests

Introduction

For any test to play a valuable role in the decision-making process, it has to be matched to the abilities and competencies required by the job role. The first part of this section provides an overview of how to identify appropriate tests and introduces the facilities in the PfS-Reasoning Tests series that allow the most suitable level of each test to be selected.

Good administration, whether the tests are being taken in pencil-and-paper format or using a computer, is the key to achieving reliable and valid test results. When administering the test in person, a well-defined procedure is to be followed. However, computer administration offers test takers the opportunity to complete tests in their own time, at a location of their choosing, without an administrator being present. Under these conditions the administration procedure may not be as closely controlled, but it is still possible for clear guidelines to be established. The second part of this section outlines the procedure for supervised test administration and goes on to offer guidelines for organisations on how to develop procedures for unsupervised testing.

Selecting appropriate tests

The information provided by the PfS-Reasoning Tests should be valuable in the decision-making process. To make sure this is the case, the abilities being assessed by the tests must relate to core job competencies. The starting point for any selection or development process must be a detailed job analysis focussing on the competencies and personal characteristics that employees need in order to perform successfully. As job roles and organisational structures become ever more fluid, identifying and assessing the competencies needed for those who work in these changing environments can also help organisations plan for future development and growth.

It is important to remember that reasoning tests can provide valuable information, but are rarely sufficient on their own. Tests should be seen as only one part of an overall assessment package. As with any form of assessment, both their strengths and weaknesses need to be acknowledged. Through drawing on the strengths of a variety of assessment methods and carefully integrating the information from them, it is possible to reach far more valid and defensible decisions that are also more likely to be viewed as fair within the framework of employment law.

In order to provide maximum information about individuals, it is important that the correct level of each Reasoning Test is selected. If tests are not at the correct level for the group in question, their ability to differentiate between people is lowered and they may have a de-motivating effect on those who take them. It is important to recognise that selecting more difficult tests will not result in a raising of standards within an organisation.

© 2003 – 2013 Team Focus Limited

Tests give most information when scores are spread around the mid-point of the distribution; if they are too easy or too hard, scores will be more bunched together so making it difficult to reliably differentiate between the test takers. The availability of appropriate norm groups is another factor in determining test selection and also indicates for which ability levels or groups tests are suitable.

Currently, there are four levels of each of closed Reasoning Tests and two levels of the open Reasoning Tests (referred to in the online PfS assessment system as 'Reasoning Skills Tests' to differentiate them from the closed tests). In addition, there is also the Combined Reasoning Test which includes verbal, numerical and abstract items in a single test. For this Combined Reasoning Test there is just one level. Each level has been developed to correspond to a broad ability band, as shown in **Table 1**. These bands should be considered as a starting point for test selection.

Reasoning Test level	Reasoning Skills Test level	Approximate educational level of the norm group
Level 1	Level 1 and Combined Reasoning Skills Test	This covers the top 95% of the population and is broadly representative of the general population.
Level 2		This covers the top 60% of the population and is broadly representative of people who have studied for A/AS Levels, GNVQ Advanced, NVQ Level 3 and professional qualifications below degree level
Level 3	Level 2	This covers the top 40% of the population and is broadly representative of the population who study for a degree at a British University or for the BTEC Higher National Diploma/Certificate, NVQ Level 4 and other professional qualifications at degree level
Level 4		This covers the top 10% of the population and is broadly representative of the population who have a postgraduate qualification, NVQ Level 5 and other professional qualifications above degree level

Table 1: Correspondence between the PfS-Reasoning Tests and level of ability as indicated by the level of educational attainment

When using the tests with well-defined groups, such as A-level students or those just about to graduate, **Table 1** should be adequate for appropriate test selection. Deciding on the appropriate level is more difficult when the tests are being used for a less homogenous group, particularly when some people may have considerable work experience but limited academic qualifications. When the most suitable test level is not immediately apparent, users may consider identifying the appropriate level by using a 'locator' test. A description of how to use a locator test is given below.

© 2003 – 2013 Team Focus Limited

Using the PfS- Reasoning Tests as locator tests

To identify the appropriate level of the PfS-Reasoning Tests, it is possible to use one of the tests as a locator test. Either the paper- or computer-based tests can be used in this way, but it is more efficient and flexible to use the computer-based tests. The locator test approach is possible because of the common format of the PfS-Reasoning Tests, and simplified by the time-efficient nature of the tests and computer-based scoring and reporting.

By administering locator tests to current employees, it is possible to establish a mean level of reasoning abilities within specific groups. This information can then be used to select the most appropriate level test for the position in question. It is suggested that Level 2 of the closed PfS-Reasoning Tests are used as the locator tests, as these should not be found too difficult by employees and will give an appropriate indication of which of the four levels is most suitable. If only one of the three test types – Verbal, Numerical or Abstract – will eventually be used, then this one should be the locator test. If two of three tests are being used, it is suggested that the test with the highest level of face validity for the role in question is used so as to get most buy-in and motivation from the volunteers being asked to complete it.

Recent research using the Level 2 tests as baseline measures has also provided a way of equating test scores across different levels of test. This can be useful if estimates of test performance on higher level tests are required. See Section Four for further details.

The locator tests provide a method for identifying reasoning abilities in current employees. It is recognised that many employees will have developed their skills in key job areas since being employed through training programmes, job experience or a combination of the two. Although it is not possible to determine the actual extent of skill growth, allowance for this is made through recommended test levels being adjusted slightly downward where borderline scores occur.

As with any test administration, it is important that good practice is followed for the administration of the locator tests if they are to provide valid information. Whilst it is not necessary to conduct a formal group administration, the following stages are recommended:

- Identify the group of current employees to take the locator test. As far as possible, this group should be at the same level as those being selected and working in the same job role.

- Ideally between 10 and 20 people should take the locator test. Asking for volunteers is likely to result in a sample that is more comfortable, and probably capable, with reasoning tests. It is not best practice to make it compulsory for employees to take the locator test. Hence, there are two possible ways in which groups can be identified:

  – A random sample from all suitable employees can be taken.

© 2003 – 2013 Team Focus Limited

- If there is a need to raise the skill level in the specific reasoning area, or if tests are to be used for development purposes, a sample can be taken from employees who are known to perform well in their job roles. However, it is important to ensure that the test level identified is not too high, meaning that tests will not adequately discriminate between potential employees. Excluding the bottom 20 or 25% of performers in the area assessed by the Reasoning Test may be an appropriate cut-off for identifying a high-performing group.

- Selected employees should be contacted, either by letter or email, requesting their participation in the locator testing. The purpose of the testing should be made clear, as should the confidential and anonymous nature of the test results. Guidelines should be given about taking the tests, for example that they should be taken in a quiet environment, free from disturbances. Test takers should also be told how long they should allow for completing the test. Clear information about the locator test is important to ensure that employees 'buy in' to the testing process and are motivated to perform at their best.

- The website address and passwords can be sent to employees at the same time as their participation in the testing is requested. Alternatively, employees can be asked whether they are willing to participate and the website address and passwords sent to those who agree. If paper-based tests are being used, details of when and where the test session will take place should be sent out. Administration procedures should follow the guidance given later in this section.

- If employees are not to receive individual results from the locator test, it is important to acknowledge their contribution to the process and to thank them for their participation.

When sufficient data has been collected, the mean of the raw test scores should be calculated. As mean scores can be affected by individual values that are far from the majority of scores ('outliers'), data should be visually inspected to check whether there are any extreme values. If there are any scores that are 6 or more raw score points below all others, it is recommended that these are removed before the mean is calculated. Lower scores may be particularly a problem due to the motivation of some test takers. Higher scores should not be removed as these will reflect high levels of ability in the sample.

© 2003 – 2013 Team Focus Limited

Table 2 shows the recommended level of the Verbal, Numerical and Abstract Tests, according to mean locator test score. Note that these scores are based on the percentiles from the Level 2 test, using the norms given on pages 83, 87 and 91 for Verbal, Numerical and Abstract respectively

Locator test percentile score	Recommended PfS test level	
	Reasoning Tests (closed)	Reasoning Skills Tests (open)
1-35	Level 1	Level 1 and Combined Test
36-70	Level 2	
71-90	Level 3	Level 2
91-99	Level 4	

Table 2: Appropriate Verbal, Numerical and Abstract test levels according to locator test percentile scores

Administering paper-based tests

Overview of administration

For a test session to be fair and to fulfil the purpose for which it was designed, it is important that it is run efficiently and smoothly. The way a test session is delivered can potentially affect the anxiety and performance of the test takers, their impression of the organisation and their motivation to perform well. The aim is for the administrator to be personable, efficient and clear when giving the test instructions.

This part of the User's Guide gives full instructions on how to prepare for administering the PfS-Reasoning Tests. In addition, there is a separate card of Administration Instructions for each test, which sets out the exact procedure to follow for the test session. Administrators are advised to prepare using this User's Guide in conjunction with the Administration Instructions and then, in the test session itself, just to use the Administration Instructions and any personal notes they have made.

For each test, administrators need to familiarise themselves with the:

- Question Booklet
- Answers to the example and practice questions
- Answer Sheet
- Administration Instructions card
- Test Log.

Before administering any of the tests, administrators should take the tests themselves – this is the best way to understand what is required. The procedure set out on the Administration Instructions card should be practised and administrators should make sure that they fully understand the solutions to the example and practice questions (full explanations to the practice questions are given in **Appendix One**).

© 2003 – 2013 Team Focus Limited

The PfS-Reasoning Tests can be administered in any order although the most usual is:

- Verbal
- Numerical
- Abstract.

Planning the test session

The test room needs to be suitably heated and ventilated (with blinds if glaring sunlight is likely to be a problem) for the number of people taking the tests and for the length of the test session. The room should be free from noise and interruption, as any disturbances can affect test takers' performance. There should be space between each test taker's desk so that test takers cannot see others' papers and the administrator can walk around.

If the tests are to be taken as part of an assessment day, remember that performance tends to deteriorate towards the end of a long day. If a number of test sessions are being planned, those who take the tests towards the end of the day may be disadvantaged. It is recommended that test takers can take any two of the Reasoning Tests without a break being needed between them. If all three tests are being administered, there needs to be a break, normally between the second and third tests. A break of at least ten minutes is recommended.

If more than 15 candidates are attending the test session, it is advisable for the administrator to have a colleague to assist with the administration. The efficiency of the session will be improved if a second person can check that test takers have grasped the practice questions and format of the Answer Sheets, and to assist with administration generally. Some preparation is also necessary for this role, particularly familiarisation with the Question Booklets, Answer Sheets and explanations to the practice items.

Test takers should be notified of the date, time and location of the test session and told which test(s) they will be taking. The Test Taker's Guide can be sent out at this stage, to help candidates prepare. The Test Taker's Guide is available online, allowing test takers to be contacted by email, if appropriate. This method may be particularly useful if test takers will be completing the computer-based tests. At this point, it is good practice to inform candidates why they have been asked to take the tests, how the results will be used in the selection procedure, how they will receive feedback about their performance and to explain the organisation's policy on the confidentiality of test results. The Test Taker's Guide can be accessed from the following link to the PfS website:

www.profilingforsuccess.com/about/documents/test_takers_guide.pdf

© 2003 – 2013 Team Focus Limited

When test takers are notified about the session, it is essential that they are also asked to contact the administrator or other appropriate person, if they have any disabilities that will affect their ability to complete the tests and to specify what accommodation needs to be made for them to complete the tests. Under the Disability Discrimination Act (1995; 2005), test users are obliged to make changes to assessment procedures so that people with disabilities are not disadvantaged at any stage of the selection process. By obtaining information about any special needs well in advance of the test session, organisations can make the necessary adaptations to the testing session and have time to seek further advice if necessary. Further information on assessing people with disabilities can be found on the PfS website as:

www.profilingforsuccess.com/about/documents/Assessing_People_with_Disabilities.pdf

Materials

Before the testing session, ensure that there are the correct number of Question Booklets and Answer Sheets. Question Booklets should be checked to make sure that they have not been marked. Marks should be erased if possible, or replacement books obtained. The Test Log has been developed to help administrators prepare for the testing session – it contains a checklist of the materials needed and other arrangements that have to be made. It also allows administrators to record the room layout, any unusual occurrences during the test session and to summarise the test scores of a group of test takers. It is a useful document to keep for later review sessions or if any challenges are made to the test results or decisions that the results feed into.

Each test taker needs:

- a Question Booklet
- an Answer Sheet
- two ball-point pens or pencils (pencils need to be sharp to clearly mark the carbonated answer sheet)
- two sheets of paper for rough working
- a candidate ID number (if applicable).

The administrator needs:

- a copy of the appropriate Question Booklet and Answer Sheet
- the appropriate Administration Instructions card
- a Test Log
- spare pens/pencils
- spare rough paper
- two stopwatches or watches with a second hand
- explanations to the practice questions if not fully familiar with them.

© 2003 – 2013 Team Focus Limited

There is space for test takers to record personal information on the Answer Sheets. Not all of this information may be needed, so administrators should make sure they are clear about what information is required and ask test takers to complete only what is necessary.

The test session

A notice to the effect of 'Testing in progress – Do not disturb' should be displayed on the door of the test room. Ensure that chairs and desks are correctly positioned. Place two pens or pencils (these need to be sharp to clearly mark the carbonated Answer Sheet), two sheets of paper for rough working, and ID numbers (if applicable) on each test taker's desk. Do not issue the Question Booklets and Answer Sheets at this stage.

If ID numbers are being used but have not already been allocated to test takers, allocate these outside the test room, then ask test takers to enter the room and find the corresponding desk. Otherwise, invite test takers into the test room and direct them where to sit.

Stage 1: Informal introduction

When all test takers are seated, the administrator should give the informal introduction to the test session. This needs to be prepared in advance to include the points given below, but should be delivered informally, in the administrator's own words. The aim here is to explain clearly to the test takers what to expect and to give them some background information about the tests and why they are being used. This will help to reduce anxiety levels and create a calm test setting. The administrator should aim for a relaxed, personable, efficient tone, beginning by thanking the test takers for attending.

The important points to include in the informal introduction are:

- Introduce the administrator (and any colleagues, if appropriate) giving their position in the company.
- The programme for the test session including: the timing, which tests will be taken, how long each test will last and the timing of any breaks (use a flipchart to show the programme if it is at all complex).
- Why the organisation is using the tests, who will see the results and how these will be used in the selection process. Explain what will happen to the test results and how they will be recorded and stored, emphasising confidentiality and accessibility in accordance with the Data Protection Act.
- Check comfort levels and whether anyone needs the cloakroom, as test takers will be asked not to leave the room once the tests have begun.
- Explain how test takers will receive feedback about their performance.
- Tell test takers that they will be given full instructions before each test, will be able to see examples, try practice questions and ask questions before the test begins, to make sure they fully understand what they have to do. Confirm that all tests will be timed.

© 2003 – 2013 Team Focus Limited

- Ask the test takers if they have any questions so far, and address these.

At the end of the informal introductory talk, test takers should be told that from this point the tests will be administered according to a set procedure and that the instructions will be read from a card, to ensure that all candidates receive exactly the same instructions. The administrator should now turn to point 4 on the appropriate Administration Instructions card and follow the exact procedure and wording given.

Stage 2: Formal testing procedure

Begin the formal testing procedure at point 4 on the relevant Administration Instructions card. It is important to follow the given procedure and wording exactly, to ensure that the instructions are the same and therefore fair and consistently administered to all test takers.

On each Administration Instructions card, the text in the shaded boxes should be read out verbatim to test takers. The text outside of the shaded boxes contains instructions for the administrator.

Use the Test Log to note the number of Question Booklets and Answer Sheets distributed and collected in, to ensure that none go astray. The start and finish time of each test should also be recorded on the Test Log. There is also room on the Test Log to record anything that occurs during the test, relating to individuals (e.g. the need for replacement pens or to leave the test room) or to the group as a whole (e.g. fire alarm or other disturbance). This information can be important later – for example, when comparing the performance of groups from different test sessions, or if an individual queries the basis of his or her selection or other decision based on test performance.

At the end of the test, collect in the Question Booklets and Answer Sheets while the test takers are still seated, ensuring while doing this that each test-taker has entered any required biographical details on the answer sheet and have indicated which level of the test they have taken. If several tests are being administered, replace any pens/pencils and rough paper, if necessary. Start the procedure for the next test from point 4 on the relevant Administration Instructions card. At the end of the session, thank test takers for attending and explain what they should do next.

Administering computer-based tests

Computer-based testing offers users far greater flexibility than paper-based tests. It also benefits from automated scoring and the ability to produce full reports almost instantly. Procedures for administering computer-based testing, particularly testing over the internet, are not as well-established as for paper-based testing. This part of the User's Guide discusses some of the options for computer-based testing. It does not set out to prescribe a process, but introduces the issues that need to be considered and makes some recommendations, so that users can formulate their own policies in this area. Administering computer-based tests under supervised and

© 2003 – 2013 Team Focus Limited

unsupervised conditions will now be considered. The technical requirements for the computer-based tests are also described.

Supervised assessment

Computer-based tests can be used as an alternative to paper-based tests. Here, test takers, either as individuals or in groups, complete the tests under supervised conditions as they would paper-based tests. The formal test instructions, example and practice items are given on-screen and so do not need to be read from the Administration Instructions card. An appropriate approach to test administration in this situation would be as follows:

- Check that the room and computers are set up appropriately.
- Invite test takers into the testing room and direct them where to sit.
- Ask test takers not to touch the computers until they are told to do so.
- Give the informal introduction as for paper-based tests (see page 15), but tell the test takers that they will be taking the test on computer.
- At the end of the informal introduction, ask if there are any questions.
- Direct test takers to the PfS website and follow the appropriate link to take a test, then give them the Client code, Access code and Password to enter when prompted. Alternatively, prior to the beginning of the testing session, ensure that the PfS website has already been accessed on each computer and the entry codes entered in order that the PfS assessment facility is already displayed on screen when candidates take their places at their computers.
- Tell test takers that the computer will prompt them to enter their personal information before giving them the test instructions and practice and example items.
- Test takers should be allowed to work through the instructions at their own pace and begin the test when they are ready.
- Explain that if they have any questions or experience any difficulties during the test, they should raise their hand.

Test takers will finish the tests at slightly different times using this approach, as not everyone will work through the instructions at the same pace. If this approach is taken, administrators should decide whether to ask test takers to remain seated until everyone completes the test or whether they can leave the room when they have finished. This is likely to depend on the number of people being tested and the room set-up (i.e. how easily people can leave the room without disturbing others).

Alternatively, test takers can be asked to work through the instructions, practice and example items, and then wait until everyone is ready to begin. When everyone is ready, the administrator should ask test takers to start. Everyone will finish the testing session at the same time if this approach is used, thus eliminating the possibility of test takers who have been slower to work through the instructions being disturbed by others leaving the room.

© 2003 – 2013 Team Focus Limited

If two of the Reasoning Tests are being taken, test takers can be instructed to move on to the second test when they have completed the first. As with the paper-based tests, if all three tests are being used, it is recommended that test takers are allowed a break between the second and third tests.

Finally, it should be noted that the tests which will be displayed on the screen when test-takers enter the PfS assessment area on the PfS web site will depend on the 'Access Code' which has been used to log in to the system. Administrators should therefore ensure that they have set up an Access Code which includes only the appropriate tests and test levels which they wish to be presented. A discussion of access codes is beyond the scope of this manual, though detailed information will be provided by Team Focus to users of the PfS online assessment system.

Unsupervised assessment

The internet offers the potential to exploit the benefits of testing in new ways, but takes users into the less familiar territory of unsupervised assessment. There are many issues with unsupervised assessment: access to technology, fairness and the authenticity of test results being paramount. Despite the need to address these issues, the benefits of internet-based testing are many. Particularly notable are its efficiency and the opportunity to gather additional information to feed into the early stages of the decision-making process.

When planning an unsupervised testing session, administrators need to consider the target group and their likely access to technology. Certain groups (e.g. university students or those already working for an organisation) may have greater access to the necessary technology than others (e.g. people returning to work). Where it is anticipated that a number of potential test takers may not have access to the necessary technology, it may be advisable not to use internet testing unless other appropriate arrangements can be made. For example, it may be possible to direct test takers to places such as libraries, careers centres or an organisation's regional offices where they can take the PfS-Reasoning Tests under appropriate conditions.

Access to the necessary technology is also related to issues of fairness. If completing internet-based assessments is made a compulsory part of an application process, this may bias the process against those who do not have easy access to the necessary technology. In some cases it could also constitute deliberate discrimination and so be unlawful. Although many organisations use online application procedures, alternatives to these should be put in place (e.g. a paper-based test session available on request). Organisations may have to accept that, in some cases, test results will not be available for all applicants.

A major question with any unsupervised testing session concerns the authenticity of results. As the tests are unsupervised, there is no way of telling who has actually completed the tests or whether the intended test taker has received assistance. If the PfS Reasoning Tests are being used for development purposes or careers guidance, authenticity should be less of an issue. It is during selection that issues around authenticity are most critical.

© 2003 – 2013 Team Focus Limited

One significant advantage of internet-based testing, as mentioned above, is that psychometric tests can be used early in a selection procedure, possibly at the same time application forms are completed. If used as part of a selection decision, it is essential to be confident that the test results are indeed the work of the applicant.

Ensuring the validity of test results requires that test takers are monitored during the test session. This removes many of the advantages of internet-based testing, so it is important to encourage honesty in test takers. One way in which this can be done is to position the tests as offering potential applicants valid feedback on their abilities and the demands of the job. This would imply on the one hand, suggesting to low scorers that the job may not be well matched to their abilities, and so would be unsatisfying for them and, on the other hand, confirming to higher scorers that they appear to have the necessary basic abilities required by the job. If test scores are used to make decisions at an early stage of an application process, it may be prudent to give them a lower weighting than normal and to set lower standards of performance.

The validity of test scores is more of an issue with high scorers. One approach to dissuade people from obtaining assistance with the tests is to view them as a 'taster' to the next stage of selection where further testing will take place under more controlled conditions. If test takers know that they will have to take a similar test under supervised conditions if they proceed to the next stage of the selection process, they may be less inclined to seek assistance with the unsupervised tests. In these circumstances it may be appropriate to initially use the open versions of the Reasoning Tests, then follow these up with the closed versions under supervised conditions if it is deemed necessary to verify results.

All the issues discussed above need to be considered when undertaking unsupervised, internet assessment. Despite this, in many ways, the actual test procedure is not that different from supervised administration. The main stages of the test process remain the same, although as it is not possible to give an informal introduction to the test session, the initial contact with test takers is very important. The contact letter, email or telephone conversation should include:

- why they are being asked to take the tests.
- what tests they have to take.
- how the results will be used.
- how they will receive feedback on their test results and who will have access to them.
- the hardware/software requirements of the tests.
- appropriate conditions for taking the tests (how long they should allow, the need for a quiet room, free from disturbances).
- how to access the testing site (website address and passwords).
- when the tests should be completed.
- either a copy of, or web link to, the Test Taker's Guide, recommending that this is used to help prepare for taking the tests.
- what will happen when the tests have been completed.
- the details of who should be contacted in case of queries or difficulties.

© 2003 – 2013 Team Focus Limited

Particularly important under unsupervised test conditions will be the information on why the tests are being used. As discussed above, positioning the tests as providing applicants with an insight into their own suitability for the job can help to encourage honesty and acceptance of the remote testing experience when used for selection. If applicants who proceed to the next stage will have to take further tests, this should also be stated, again to encourage honesty.

Once test results have been received, an opportunity should be made for test takers to discuss their results (see **Section Three**).

Technical requirements for computer-based tests

If internet testing is being considered, the issue of access to technology needs to be addressed. Although the majority of people now have access to computers, it should not be assumed that this is the case for everyone. It also needs to be recognised that conditions should be conducive to completing a timed test; some computers that are accessible to the public may be in noisy environments and where test takers are liable to disruption.

To make the PfS-Reasoning Tests widely accessible, the system has been designed to make minimal demands on technology. The system will work on any internet-ready computer. The preferred browser is Internet Explorer version 5 or later with Adobe Flash® version 5 or later installed. The minimum screen resolution needed is 800 x 600 though a resolution of 1024 by 768 is recommended. Virtually all modern desktop computers and most modern laptop computers will meet the specifications needed to run the tests. Tests are accessed over the internet. As the whole test is downloaded before the test begins, timing for the test is unaffected by the speed of the internet connection.

It is not necessary for the internet connection to be maintained once a test has been downloaded. However, the internet connection does have to be active when the test results are submitted. Information about the need for test takers to be actively connected to the internet for their test results to be recorded is displayed at the end of the test.

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Section Three: Scoring and review of test results

Overview of scoring and test scores

The primary purpose of testing is to obtain a test score or mark which says something about the test taker's ability. To understand any test score however, it needs to be put into context. For example, if a person has answered 26 out of a possible 30 questions correctly, this appears initially to be a good score. But if another group of other test takers all score between 27 and 30 on the same test, is 26 still a good score?

The purpose of this example is to highlight that simple test scores cannot be considered 'good' or 'poor', without knowing more about how people generally perform on the test. Test scores are put into context by comparing them with the scores of a large group of people who have previously taken the test. This group is known as the 'norm group' and the tables that allow individual scores to be compared to those from the norm group are called 'norm tables'. The norm tables for the PfS-Reasoning Tests are in **Appendix Three**. The types of scores given in the PfS-Reasoning Tests norm tables are described below.

- Raw score – The raw score is the number of marks a test taker achieves on a test. For the Verbal, Numerical and Abstract Reasoning Tests, one mark is given for each question that is answered correctly. Therefore, the raw score is the number of questions answered correctly.

- Percentile – Percentiles describe the proportion of the norm group a test taker has scored the same as or better than. For example, a percentile of 65 means that the person has scored as well as or better than 65 percent of the norm group. Because percentiles are quite easy to understand, they can be particularly useful when communicating information to test takers or people who are unfamiliar with psychometric testing.

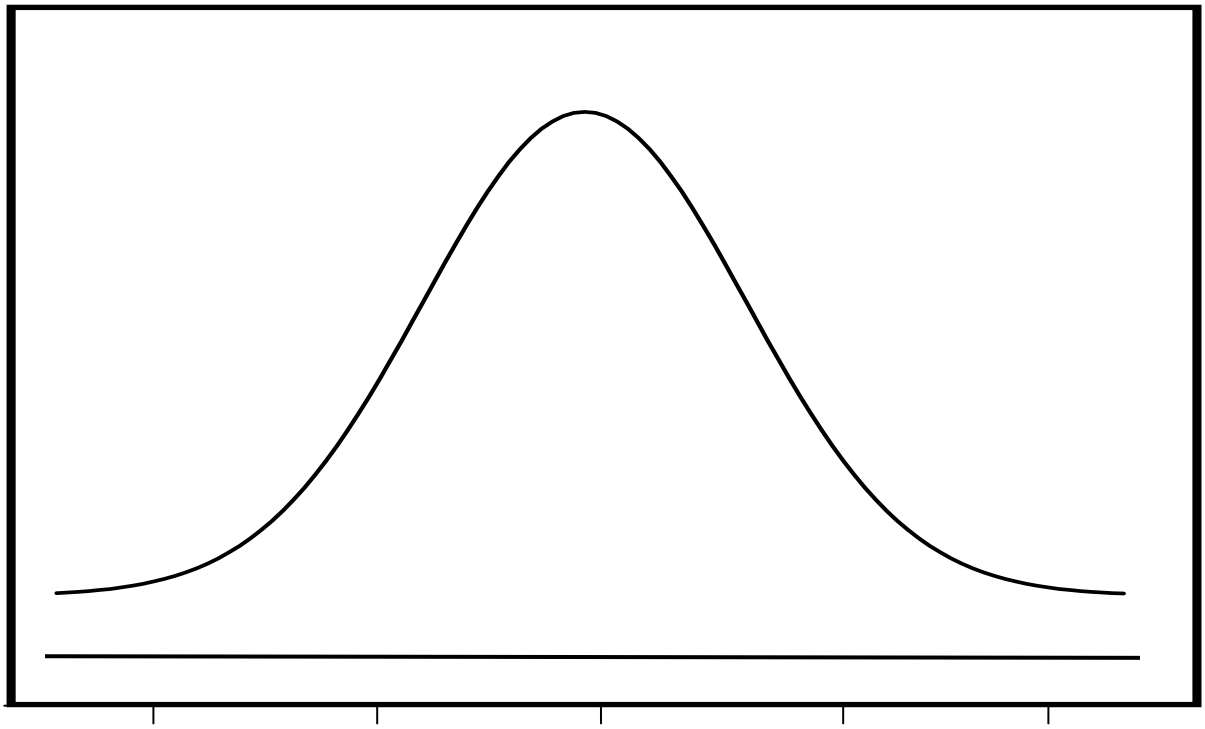
- T-score – T-scores are a transformation of the raw scores onto a scale which is approximately normally distributed (that is, a bell-shaped distribution with no long tails). This transformation is necessary as raw score distributions are often skewed, with more scores towards the higher or lower end of the distribution. T-scores have a mean of 50 and a standard deviation (SD; an indication of the spread of scores) of 10. The main advantage of using a scaled score such as T-scores, is that they allow performance on different tests to be directly compared.

- T-score confidence band – All test scores contain a degree of error, as no test can give a perfect indication of a person's ability. This error can be quantified and described as a range within which a person's 'true score' is likely to fall. The norm tables give 68% and 80% confidence bands for each T-score. These confidence bands indicate the range in T-scores between which it is 68% or 80% certain that a person's true score will lie. For a more detailed discussion of test error and confidence bands see the section on Reliability (pages 41 - 47).

© 2003 – 2013 Team Focus Limited

The relationship between percentiles, T-scores and the normal distribution curve are shown in **Figure 3**.

Figure 3: The normal distribution curve, Z-score, T-score and percentile scales



Z-score	-2	-1	0	1	2
T-score	30	40	50	60	70
Percentile	2	16	50	84	98

Qualitative analysis of results

In addition to providing a single test score, the PfS-Reasoning Tests combine two further test statistics to produce a more qualitative assessment of each test taker's performance; the number of questions the test taker has attempted and the proportion of questions attempted that have been answered correctly. Both of these values can be compared to norm tables in the same way as the raw score and are classified as being 'above average', 'average' or 'below average'. This process results in a three by three matrix, describing the test taker's speed of working and accuracy as shown in **Table 3** overleaf.

© 2003 – 2013 Team Focus Limited

		Accuracy (proportion of questions attempted answered correctly)		
		Below average	Average	Above average
Speed (number of questions attempted)	Below average	Slow and inaccurate	Slow and moderately accurate	Slow and accurate
	Average	Average speed and inaccurate	Average speed and accuracy	Average speed and accurate
	Above average	Fast and inaccurate	Fast and moderately accurate	Fast and accurate

Table 3: Summary descriptions for combinations of speed of working and accuracy

The analysis of speed and accuracy has been developed to help both test takers and users gain a fuller understanding of the Reasoning Test results. The nine summaries of test performance given in **Table 3** have been expanded into more detailed descriptions of performance, interview prompts and development suggestions. These descriptions are included in the full reports generated by computer-based tests or from the test scoring facility on the Profiling for Success website. Summary reports provide an overview the test taker's speed and accuracy but do not include full interview prompts or development suggestions.

When using the extended analyses in the full versions of the reports, it needs to be recognised that these reports offer a range of possible reasons for a person's performance. These are offered as ideas and prompts that can be used during a review session to explore test performance in more detail. Every effort has been made to make these reports comprehensive, although they should not be seen as exhaustive. Further, the reports attempt to reflect the test taker's ability, approach to the test and mental processes, but may be limited in some cases as the descriptions can be extrapolated only from the test taker's responses.

Scoring paper-based tests

Answer Sheets for the paper-based tests are made up of two sheets of paper; the top sheet where test takers mark their answers and the bottom sheet which contains the scoring key. As the top sheet is carbonated, the test taker's biographical information and answers are transferred on to the bottom sheet with the scoring key. The steps that should be followed when scoring Answer Sheets are set out below.

1. Before beginning to score an Answer Sheet, check that the test taker has completed all the necessary biographical details and that they have indicated which level of the test they have taken.

© 2003 – 2013 Team Focus Limited

2. On the right hand side of the Answer Sheet there is a perforated strip. Tear off this strip and then use a pencil or ruler to separate the top and bottom pages of the answer sheet.

3. Check that the test taker's personal details and answers to the questions have transferred clearly on to the bottom part of the Answer Sheet.

4. Count up and tick the number of times the responses given by the test taker correspond to the correct answers indicated on the bottom part of the Answer Sheet. As each correct answer is worth one mark, the total number of ticks is their 'raw score'. Enter the raw score in the box marked 'Raw score'.

5. Count up the number of questions to which the test taker has given an incorrect or ambiguous response, and add this to their raw score. This gives the number of questions that the test taker has attempted. Enter the number of questions they have given a response to in the box marked 'Number of questions attempted'.

6. Use the appropriate norm table to look up the percentile, T-score and confidence bands that correspond to the T-score. These should be entered in the appropriate boxes on the answer sheet.

7. On the reverse of the bottom part of the Answer Sheet test takers may have recorded comments about the test and the test session. This information should be available to the person conducting the review session, as it can provide useful information to discuss during the review.

Sometimes test takers make ambiguous marks on answer sheets. The following guidelines for resolving ambiguities were applied during the development of norms for the paper-based Reasoning Tests. These should be followed to ensure the validity of normative data.

- If more than one answer has been indicated to a question, and all but one answer is clearly crossed out, count this as the intended answer and score against the scoring key.
- If more than one answer has been indicated to a question, score as incorrect.
- If all answers have been crossed out, score as incorrect.
- Test takers may miss out a question but forget to leave a blank space for the question on their answer sheet. This is most apparent when a series of answers are incorrect according to the scoring key, but each indicates the correct answer for the following question. If a series of four or more answers indicate the correct answer to the following questions, it is possible that an answer has been missed out. In such cases, appropriate adjustments should be made and the questions treated as correct.

© 2003 – 2013 Team Focus Limited

Scoring computer-based tests

Computer-based tests are automatically scored when the answers are submitted at the end of the test. From the scored data a report including the raw score, percentile, T-score and confidence bands are automatically created for each test taker. An extended analysis, as described on pages 22 and 23, is also included if the full version of the report is requested. This report is sent to the email address entered by the administrator during the set-up stage of the testing process.

When setting up an assessment in the PfS 'Client Area', there is also an option for users to request reports for the test taker. Test taker's reports are versions of the main reports in a format that can be given directly to the test taker. As with the administrator's reports, full or abbreviated versions of these reports are available. If test taker's reports have been requested, these will also be sent to the email address entered by the test-taker when logging in to the PfS system.

Samples of full reports and a summary reports can be seen in **Appendix Two**.

Using the online report generator with paper-based tests

Users of the paper-based tests can also make use of the online report generator that is a standard part of the computer-based tests. The report generator requires users to be set up as clients of the Profiling for Success system, which is accessed via the internet at the following address:

www.profilingforsuccess.com

The test system will ask users to enter their Client Code and Password. Test data can then be entered through the 'direct data entry' screens. Reports will be generated on the submission of the data. For more information on this system or to set up an online PfS account please contact Team Focus.

Review of test results

Whenever psychometric tests are administered it is good practice to review the results with the test taker. The exact format of this review will depend on the purpose of assessment and how the results are to be used. Practical considerations, such as the number of people being tested and the opportunities to meet with test takers if assessment is being conducted over the internet, will also affect how the review of test results is delivered. However, regardless of the specific situation, test takers should always be given the option to discuss their results and to raise any questions they have. The following sections provide guidance on how test results can be communicated to test takers and how to conduct a review session.

© 2003 – 2013 Team Focus Limited

Communicating test results

There are three main ways in which results can be communicated to test takers. These are outlined below, along with some of the considerations around each method.

- *Face-to-face review session*. The preferred method of communicating test results is to do so in person (guidance on conducting a personal review session is given below). Face-to-face reviews have the advantages of encouraging openness and honesty, allowing reviewers greater insight into the test taker's reactions to the results and so opportunities to respond to these, and generally encourage greater interaction. The results from a Verbal, Numerical and Abstract Reasoning Test can be covered in a single review session lasting between 10 and 15 minutes, so these review sessions do not have to be time-consuming. These can be scheduled to follow testing sessions or interviews, to avoid difficulties in arranging subsequent meetings for the reviews. The option of reviewing results almost immediately after tests have been completed is possible due to the rapid scoring and report generating facilities of the PfS-Reasoning Tests, particularly when the computer-based tests are used.

- *Telephone review*. When there is no personal contact with the test taker (for example, when initial screening has been conducted over the internet and some candidates have not progressed to the next stage of assessment), telephone review sessions can be conducted. A mutually convenient time needs to be arranged between the reviewer and test taker, to ensure that both have sufficient time, free from interruptions, for the review to be conducted fully. A particular limitation of this approach is that reviewers do not have access to non-verbal cues, which can be valuable in gauging a test taker's reactions during face-to-face reviews. Under these conditions, reviewers need to be particularly aware of emotional reactions in what test takers say and may need to prompt more around how the test taker is feeling about their results than when conducting face-to-face reviews.

- *Written review*. Giving test takers purely written information on their test performance is the least preferable way of communicating results. This method essentially gives 'feedback' (test results being delivered to the test taker) as there are very limited opportunities for exploring the meaning and implications of the test results. Whenever this method is used, it is important that test takers are given the name and telephone number of a person they can contact to discuss their results.

© 2003 – 2013 Team Focus Limited

The PfS-Reasoning Tests can produce reports for both the reviewers and test takers. The test takers' reports (see **Appendix Two**), either in their full or summary versions, are suitable for giving directly to test takers as care has been taken in developing these reports to ensure that the language used is accessible and generally positive. Test takers' reports give raw scores and percentiles in written and graphical form, and include personal development suggestions. The score bands used in the administrator's reports and their relationship to T-scores and percentiles are shown in **Table 4.**

Score band used in summary report	T-score band	Percentile band
Low	36 and below	<1-10
Below average	41-37	11-30
Average	42-58	31-69
Above average	59-63	70-89
High	64 and above	90-99

Table 4: Score bands used in administrator's reports and their relationship to T-scores and percentiles

Reviewers need to consider if they want to use the test takers' reports and, if so, when they will be given to test takers. It can be useful to give the reports to test takers before the review session, allowing time for the reports to be read and test takers to think about issues they may want to raise in the review session. In principle, reports can be given to test takers during the review session or at the end of it. However, if reports are given during the review session, test takers may be distracted by the reports unless time is allowed for them to be read.

Alternatively, reviewers can use information gained through the review session to edit and tailor reports before giving them to test takers. This approach may be particularly useful in developmental contexts, when personal development suggestions and action plans can be included in the final report given to test takers.

If a report is to be edited after a review, it is suggested that the following process is used:

1. The report and any associated graphics files are saved from the email to a folder on the computer.
2. The report file can be opened by a word processing package such as Microsoft Word®. To do this it may be necessary to select the option to allow html file types to be viewed and opened.
3. The report can then be edited as a normal Word® document and saved in its original html format.

© 2003 – 2013 Team Focus Limited

Conducting a review session

The purpose of a review session, whether conducted face-to-face or via the telephone, is to ensure that the test taker clearly understands the meaning of their results, is satisfied with the assessment experience and to explore possible implications of the results. To reach these goals it is important that the review session is seen as a chance for information to be given and received by both the test taker and the reviewer, not simply for the reviewer to provide the test scores. For this process to be successful, it is vital that all reviewers have received appropriate training.

General guidelines for conducting review sessions are given below. These guidelines should be seen as identifying the main points that need to be covered and giving suggestions about the structure of the review session and appropriate questioning strategies. They do not set out to provide a set formula that must be followed. Although the guidelines below are written with face-to-face reviews in mind, they are also applicable to telephone reviews.

- As with test administration, good preparation is essential for review sessions. A suitable room, free from disturbances, should be identified. Reviewers should familiarise themselves with the individual's test results, what the test measures and how this relates to the job role, and any other relevant biographical information. Technical language should not be used during the review session, so it is useful for reviewers to prepare a simple description of what each test measures. For example, a Numerical Reasoning Test may be better described as 'an opportunity to show how comfortable you are with using numbers and numerical information to solve problems'. Reports should be sent to test takers in good time if these are being given out before the review session.

- The review session should begin with the reviewer introducing themselves and providing a brief overview of the purpose of the review session. Useful information to provide includes the approximate length of the session, issues around confidentiality and what will happen to the test results.

- Both parties need to agree on what they want to get out of the session, such as information, consequences of test performance or a way forward.

- To encourage a balanced discussion from the outset, the test taker should be brought into the review session as early as possible. This can be done through asking the test taker about their experiences of the tests immediately after the brief introduction (e.g. "How did you find the reasoning tests?" or "Tell me about your experience of taking the tests?"). Throughout the review session open questions should be used wherever possible, as this will encourage the test taker to provide more information and make the review more balanced. In a balanced review session there should be equal contributions from both the reviewer and the test taker.

© 2003 – 2013 Team Focus Limited

- If the tests were completed some time before, a reminder of these and how they fit into the selection or development process may need to be given at this stage.

- At this point it is also appropriate to explain how test results are interpreted with reference to a norm group. It is generally best to avoid the term 'norm group' as this may not be understood by all test takers and for some may imply 'normal' performance. A preferable phrase is 'comparison group', which conveys the process of comparing individual test scores to those from a wider group of people, and is more readily understood.

- The next stage involves discussion of the actual test results. It is preferable to let the test taker lead the order in which the tests are reviewed, rather than going through the tests in order. The review process can be started through questions such as "Which test did you prefer and why?" or "Which test did you find most challenging?". Once a test has been identified, the reviewer can give the test taker their score or can ask them to estimate their own performance on the test, for example "In relation to the comparison group (describe comparison group) how do you feel you performed on the (appropriate test)?".

- It is preferable to describe test scores in terms of percentiles, though it needs to be clearly communicated that percentiles refer to the proportion of the comparison group who the test taker scored as good as or better than, and not the percentage of questions they answered correctly. It may also be informative to explore the number of questions that test taker attempted and number answered correctly as this, in conjunction with the text around speed and accuracy, can be used to explore the way in which the test taker approached the test.

- Once the test taker's performance on each test has been established, their reactions to the result and its implications need to be explored. For example, questions such as "How do you feel about your result on this test?" can be used to assess emotional reaction and "What implications do you think the test results may have on your application?" or "How might your result on this test influence your choice of career?" can be used to explore the implications of test results. Although reviewers often perceive low scores as more challenging to discuss, it is important that low test scores are not 'glossed over' or dismissed. Questions such as "How far do you think the result is a fair reflection of your ability in this area?" can be very valuable. Often test takers have a reasonable insight into their abilities and low scores in some areas may not necessarily be a great source of concern; test takers often find it quite reassuring to know that they have performed at an 'average' level.

- If the computer-generated test user's reports are being used, these contain interview prompts to explore test performance and the implications of this. As these reports combine information on speed and accuracy, they offer prompts that are specifically tailored to the individual's performance and how they could improve their performance. Because of this they can be particularly valuable when using the tests for development or when the reviewer has limited experience of working with this type of assessment.

© 2003 – 2013 Team Focus Limited

- The final stage of the review process is to ask the test taker to summarise what has been discussed, to ensure clear understanding. Summaries can take the form of a brief review of the test results that highlight any strengths and weaknesses that have been identified. The implications of the test results and any development plans should also be summarised, if these have been discussed. To check that the test taker has understood what has been discussed, it can be valuable to get them to summarise what they see as the main points to have emerged from the review session, rather than this being provided by the reviewer. The reviewer should explain the next stage in the selection or development process and what will happen to the results, and inform the test taker about confidentiality. Finally, the test taker should be offered the opportunity to ask any outstanding questions and then thanked for attending the review session.

It is good practice for individual organisations to develop policies around the review of test results, as with other aspects of testing. Such policies should cover the organisation's general policy on test reviews, how reviews are conducted, confidentiality and storage of information. It is important for organisations to develop their own policies, as these will help ensure consistency of approach and application over time, and will also guard against issues of fairness and discrimination. Whilst policies may draw on the guidelines given above, ultimately reviewers should develop their own style, with which they feel comfortable, within these frameworks.

© 2003 – 2013 Team Focus Limited

Section Four: Development of the Reasoning Tests

Test formats

The development of the Verbal, Numerical and Abstract Reasoning Tests involved a number of stages. The purpose of the first stage was to define as clearly as possible, the final format of the tests. By understanding how the final tests would look and function, the test development team identified the main determinants of the item formats. The key aspects affecting item formats were identified as:

- Ability range – The tests should be suitable for a wide range of abilities; from people with average GCSE passes or equivalent, up to postgraduates and those with considerable professional experience. It was therefore necessary to identify item formats that could support questions across this ability range.

- Computer technology – From the outset it was decided that the tests should be primarily computer-based, but that many users would still want pencil-and-paper versions to be available. Item formats that could be used in both mediums were therefore needed. The test development team wanted to exploit the advantages of computer-based technology, but also recognised that this technology needed to be widely available. While the internet was seen as offering the greatest flexibility for testing, it was necessary to consider issues such as use of graphics, download times and the possible unreliability of internet connections in the design of the tests.

- Test length – Test users may often want to incorporate a broad range of assessments into their selection and development procedures, but also need to consider the time available for testing. With this in mind, the target time for each test was set at 15 to 20 minutes, depending on test level. Item formats therefore needed to be efficient (e.g. by minimising the time spent on unnecessary reading or calculations) but also needed to remain sufficiently contextualised to reflect real-life problems. The tests were therefore designed to be problem-solving tasks, presenting test takers with information and questions based on the information. To make the items 'efficient' a number of questions were related to each piece of information, or 'stem', and the length of the stems was carefully controlled.

- Answer format – To allow for the scoring of the tests to be achieved quickly and reliably, a multiple-choice format was used for each of the tests. Whilst open-response items can offer a rich source of information, this is offset by the difficulties in scoring open-response tests reliably, particularly when scoring tests by computer. The time needed for scoring and resolving ambiguities in open-response tests was also seen as unacceptable to those who use high volumes of tests.

© 2003 – 2013 Team Focus Limited

Verbal reasoning format

The Verbal Reasoning Tests consist of passages of information, with each passage being followed by a number of statements. Test takers have to judge whether each of the statements is true or false on the basis of the information in the passage, or whether there is insufficient information in the passage to determine whether the statement is true or false. In the latter case, the correct answer option is 'can't tell'.

As test takers come to the testing situation with different experiences and knowledge, the instructions state that responses to the statements should be based only on the information contained in the passages, not on any existing information that test takers have. These instructions also reflect the situation faced by many employees who have to make decisions on the basis of information presented to them. In these circumstances decision-makers are often not experts in the particular area and have to assume the information is correct, even if they do not know this for certain.

The passages of information in the Verbal Reasoning Tests cover a broad range of subjects. As far as possible, these have been selected so that they do not reflect particular occupational areas. Passages were also written to cover both emotionally neutral areas and areas in which people may hold quite strong opinions or have emotional involvement. Again, this was seen to make the Verbal Reasoning Test a valid analogy of decision-making processes, where individuals have to reason logically with both emotionally neutral and personally involving material.

Each statement has three possible answer options – true, false and can't tell – giving test takers a one-in-three or 33% chance of guessing the answer correctly. Guessing is most likely to become a factor when tests are highly speeded. The quite generous time limits and the 'not reached' figures, suggest guessing is unlikely to be a major factor for the Verbal Reasoning Tests. The proportion of true, false and can't tell answers was balanced in both the trial and final versions of the Verbal Reasoning Tests. The same answer option is never the correct answer for more than three consecutive statements.

Numerical reasoning format

The Numerical Reasoning Tests present test takers with numerical information and ask them to solve problems using that information. Some of the harder questions introduce additional information which also has to be used to solve the problem. Test takers have to select the correct answer from the list of options given with each question.

© 2003 – 2013 Team Focus Limited

Numerical items require only basic mathematical knowledge to solve them. All mathematical operations used are covered in the GCSE (Key Stage 4) mathematics syllabus, with problems reflecting how numerical information may be used in work-based contexts. Areas covered include: basic mathematical operations (+, -, x, ÷), fractions, decimals, ratios, time, powers, area, volume, weight, angles, money, approximations and basic algebra. The tests also include information presented in a variety of formats, again to reflect the skills need to extract appropriate information from a range of sources. Formats for presentation include: text, tables, bar graphs, pie charts and plans.

Each question in the numerical test is followed by five possible answer options, giving test takers just a one-in-five or 20% chance of obtaining a correct answer through guessing. The distractors (incorrect answer options) were developed to reflect the kinds of errors typically made when performing the calculations needed for each problem. The answer option 'can't tell' is included as the last option for some problems. This is included to assess test takers' ability to recognise when they have insufficient information to solve a problem. As with the Verbal Reasoning Tests, the same answer option is never the correct answer for more than three consecutive statements.

Abstract reasoning format

The Abstract Reasoning Tests are based around a categorisation task. Test takers are shown two sets of shapes, labelled 'Set A' and 'Set B'. All the shapes in Set A share a common feature or features, as do the shapes in Set B. Test takers have to identify the theme linking the shapes in each set and then decide whether further shapes belong to Set A, Set B or neither set.

The abstract classification task is based on Bongard problems (Bongard, 1972). Bongard problems were originally developed to test the ability of computer-based pattern recognition programs. In their original form these problems consisted of two sets, each containing six shapes. Computer programs had to identify the common feature(s) of the shapes in each set, but they were not required to classify further shapes.

A development of this task was chosen for the Abstract Reasoning Test as it requires a more holistic, inductive approach to problem-solving and hypothesis-generation than abstract problems involving sequences of shapes or progressions. People operating at high levels are often required to focus on different levels of detail, and to switch between these rapidly (e.g. understanding budget details and how these relate to longer-term organisational vision). These skills are assessed through the Abstract Reasoning Test, as it requires test takers to see patterns at varying levels of detail and abstraction. The lower level of the abstract test can be a particularly valuable tool for spotting potential in young people or those with less formal education, as it has minimal reliance on educational attainment and language.

© 2003 – 2013 Team Focus Limited

Test takers are required to identify whether each shape belongs to Set A, Set B or neither set. This gives three possible answer options, meaning test takers have a one-in-three chance of guessing answers correctly. As with the other tests, the proportion of items to which each option is the correct answer has been balanced. The same answer option is never the correct answer for more than four consecutive shapes.

Item writing

The test items were written by a team of people, who all had extensive experience of occupational psychology or using assessments in selection and development contexts. Detailed briefing notes were assembled for item writers, outlining the nature of the tests, the specific details of the items for each test type and giving example items. Prior to writing test items, all item writers attended a workshop which introduced them to the process of item writing and covered areas of good practice, particularly in relation to bias. This was followed by a practical session involving item writing and group review of the items produced.

After attending the workshop, item writers initially submitted a limited number of items to the test development team for review and feedback. Only after these initial items were considered satisfactory were item writers given the go-ahead to develop the test items. Throughout the item writing process, feedback was continually given to item writers to ensure standards were maintained and that adequate coverage of each area was achieved.

Pre-trialling item reviews

Before trialling of the items took place, they went through a number of review stages. As they were submitted, items were reviewed by the test development team to ensure they met the test specifications, were accurate and unambiguous, and free from bias. All items were further reviewed by an occupational psychologist who was not involved with the test development.

Following the internal item reviews, items were sent to external specialists. Verbal items were reviewed to check for clarity of language and reasoning, Numerical items were reviewed for language and mathematical accuracy (including plausibility of distractors), and Abstract items were checked for accuracy and ambiguity in solutions. Verbal items, Numerical items and instructions for all three Reasoning Tests were also reviewed by an educational psychologist who specialises in language and cultural issues to ensure they were accessible and free from gender and ethnic bias.

© 2003 – 2013 Team Focus Limited

Trialling

The majority of the trialling was computer-based, as this allowed the tests to be accessible to a wide range of organisations and individuals and reduced the possibility of errors in data collection and transfer. Computer-based trialling also allowed timing data to be collected for each test item. An analysis of timing data was included in the initial item analyses and contributed to the final selection of items, so helping to make the tests efficient in terms of time.

Each trial test lasted between 30 and 35 minutes. The trial tests were designed to be less speeded than the final tests, so that sufficient data could be collected on items later in the tests. However, a timing element was included to create realistic test conditions for the trial tests, as this was needed for accurate data for analysis and item selection. The trialling design involved placing common items in adjacent levels of each test to allow linking and the substitution of items between test levels as necessary.

The trialling software collected biographical information on each test taker. Information on age, gender, educational qualifications and ethnicity were collected for all people who took part in the trialling. Further information was also collected from specific samples as appropriate (e.g. current course of study and predicted degree grades for graduate samples). In total, almost 2000 people participated in the trialling of the items for the closed tests between January 2002 and July 2002. Trialling for the open tests was conducted between October 2002 and February 2003. Approximately 3000 people took part in this exercise.

Item analysis

For trialling, the timed part of the tests lasted between 30 and 35 minutes and each had approximately double the number of items of the final tests. Once sufficient trialling data had been collected, each test went through a series of analyses to identify items that were not functioning satisfactorily.

Item analyses were conducted to identify the facility and discrimination for each item. The facility indicates the proportion of people who attempted the item who answered it correctly, effectively indicating the difficulty of the item. The discrimination is the point biserial correlation between the score on the item (1 for correct, 0 for incorrect) and total test score excluding that item. This statistic indicates the degree to which each item is able to distinguish between people who obtained high overall test scores and those who obtained lower scores.

As each of the tests uses a multiple-choice format, ‘distractor’ analyses were conducted on the incorrect items. Essentially, these are the same as discrimination analyses, but examine the link between each of the incorrect answer options and total test score. If items are functioning well, people who choose incorrect answers should get lower overall test scores. If this is not the case, items may be ambiguous, leading strong test takers to choose incorrect answer options.

© 2003 – 2013 Team Focus Limited

The time taken by test takers to answer each question was also recorded, and the mean times for each of the questions calculated. This analysis identified items that were answered particularly quickly or slowly on average. Timing information was used primarily to ensure that the items selected for the final versions of each test obtained maximum information within the times allowed for each test. This analysis also complemented the item analysis, suggesting where items may be too easy (very rapid response times) or ambiguous (slow response times).

Each of the trial tests was subjected to a bias analysis to ensure that selected items were not found disproportionately easy or hard by different groups of test takers. Comparisons were made between males and females, and 'whites' and 'non-whites'. Items displaying significant levels of bias were excluded, or were included but balanced with an item showing an equal degree of bias for the opposite group.
Following the item analysis, tests were assembled for standardisation. The time allowed and number of items in each of the standardisation tests are given in **Table 5**.

		Closed Reasoning Tests		
		Verbal	Numerical	Abstract
Level 1	Time allowed	12	12	10
	Number of items	32	28	50
Level 2	Time allowed	12	12	10
	Number of items	32	28	50
Level 3	Time allowed	15	15	12
	Number of items	40	36	60
Level 4	Time allowed	15	15	12
	Number of items	40	36	60
		Open Reasoning Tests		
		Verbal	Numerical	Abstract
Level 1	Time allowed	15	15	12
	Number of items	44	40	70
Combined test	Time allowed	10	10	7
	Number of items	24	20	35
Level 2	Time allowed	20	20	15
	Number of items	60	48	75

Table 5: Timings and number of items in each of the PfS-Reasoning Tests

© 2003 – 2013 Team Focus Limited

Item Response Scaling of Tests

Additional analysis, using the closed versions of the Verbal, Numerical and Abstract tests, has allowed for the construction of a common comparison scale, and has also established that different levels of the tests are measuring different levels of ability.

Producing a common scale involved administering different levels of tests to the same sample of people. This took place over a number of years, with the following test combinations:

	Sample sizes		
	Verbal	Numerical	Abstract
Closed Level 1 with Closed Level 2	1008	1773	768
Closed Level 2 with Closed Level 3	777	887	757
Closed Level 3 with Closed Level 4	498	930	210
Open Level 2 with Closed level 2	1547	1293	807

Developing a common scale

All these data were analysed using Item Response Theory (IRT). As the name implies this involves looking at the statistical performance of each item (question) rather than examining tests at the level of the complete test[1]. It is used to derive an estimate of an individual's ability in terms of the known parameters of each individual item in a test. Using this approach, it is possible to equate scores obtained on one version of a given test with scores obtained on a different version, assuming that both tests are measuring the same underlying ability trait.

Practically, by estimating the difficulty of each item and the ability of each test taker, a common scale can be produced. This means that any test can generate a score based on this common scale which can then be used to estimate what a person would have obtained if they had completed a different test. This provides the methodology whereby tests can be used inter-changeably.

For the purpose of the exercise, the Level 2 versions of each of the Numerical, Verbal and Abstract Reasoning tests were considered as the base versions. The samples were as shown above, with total sample sizes varying between 1500 and 3000 records, depending on test and versions in question. In each case, the respondent took a special version of the test which contained sets of items from two of the four possible levels (1 - 4) of the test.

The outcome of the IRT analysis was a set of look-up tables which allowed a person's score on the version of the test they had actually taken to be translated to a score common across all versions of that test. This in turn demonstrated that it was

[1] Thus there is a fundamental difference between IRT and Classical Test Theory. Not only is the analysis at the level of individual items, accuracy (reliability) is estimated for each score/participant, rather than being the same for all participants; as well as the overall score being item independent; and the item parameters themselves, sample independent.

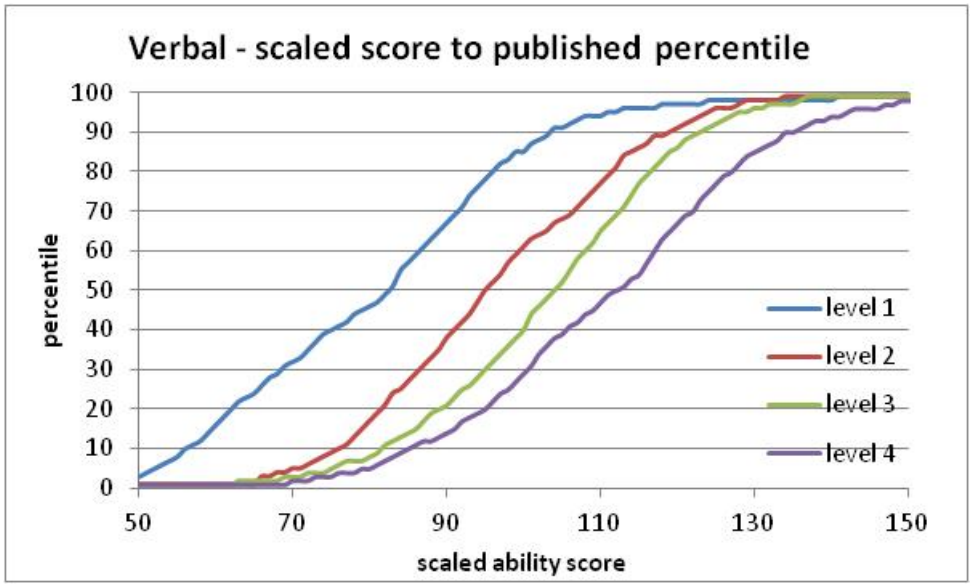
© 2003 – 2013 Team Focus Limited

possible to construct a coherent common scale that linked all levels of each of the tests. Critically it also illustrated that there is a difference between each of the tests in terms of level, i.e. the tests genuinely differ in order of difficulty from Level-1 (Easiest) to Level-4 (Hardest). The exception perhaps being the Level-1 and Level-2 Abstract tests.

This separation in difficulty can be illustrated visually using test characteristic curves. These, typically s-shaped curves, are ways of illustrating the functional relationship between the true score and the ability scale. The curves for the three tests, and the different levels of each test, are reproduced below. In particular it is worth noting that the curves for each test span the ability range from left to right - Level-1 on the left, Level-4 on the right - and they do not tend to cross. There is a clean separation between each test.

The proximity of the curves in the Abstract test is understandable since the abstract test is less highly correlated with educational level.

Figure 4



© 2003 – 2013 Team Focus Limited

Figure 5

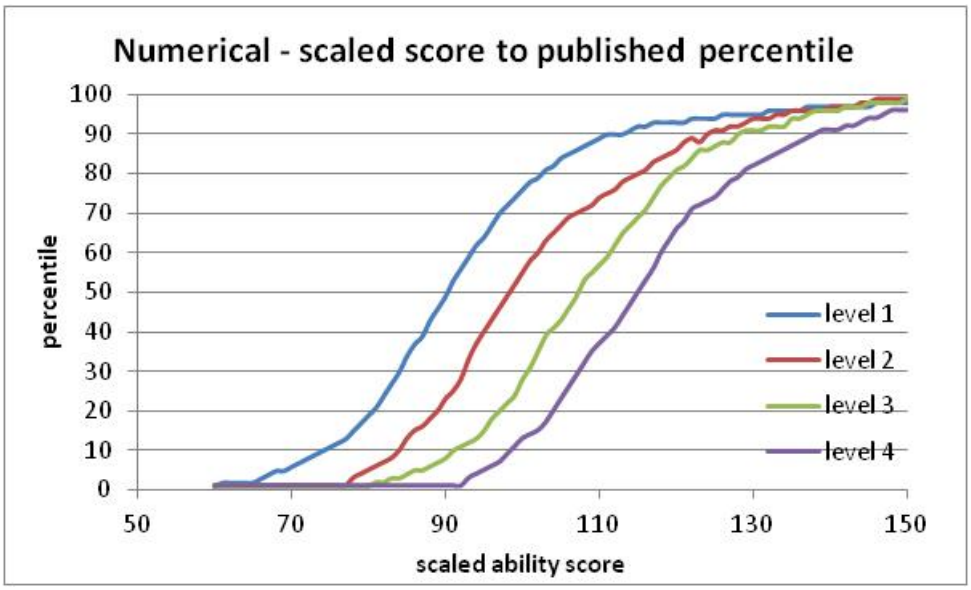
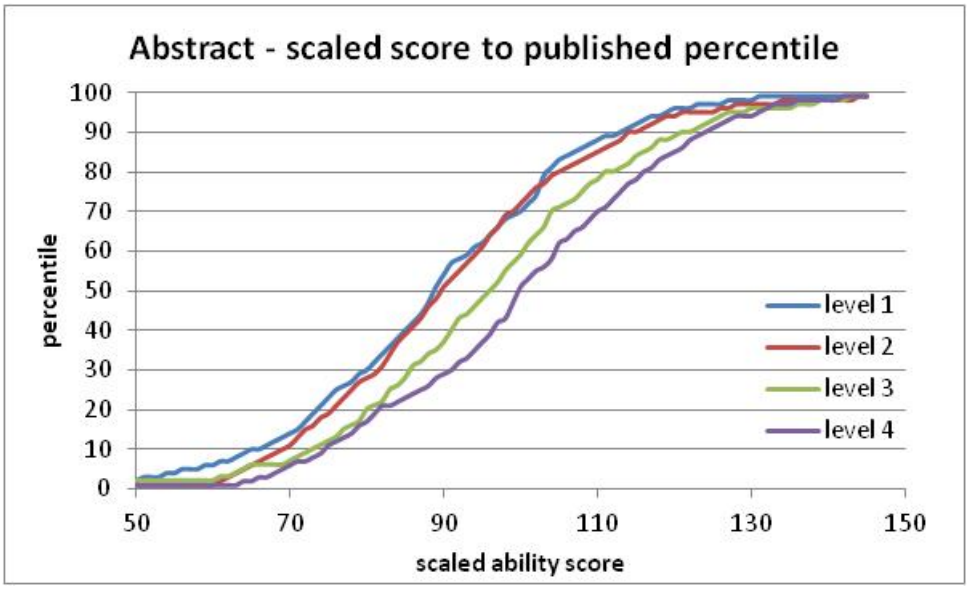


Figure 6



Note: The Team Focus IRT look-up tables are available on request.

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Section Five: Technical information

Introduction

This section of the User's Guide provides a detailed account of the technical functioning of the PfS-Reasoning Tests, covering the areas of reliability, bias and validity. The important area of reliability of measurement and the precision of test scores is explored in detail here, although the key reliability statistics – internal consistency, standard error of measurement and standard error of difference – are also summarised with each of the norm tables in **Appendix Three**.

Reliability

The concept of reliability

No test, including those in the PfS-Reasoning Tests series, gives a perfect indication of reasoning ability. Despite rigorous test development and appropriate use and administration of the tests, there will always be some degree of error in any test result. The concept of reliability is concerned with quantifying the amount of error in a test score. If the accuracy of a test score is known, then scores can be used sensitively with due regard for this error. Reliability is also important as it sets the upper limit on validity: a test cannot be more valid than it is reliable.

The need to take error into consideration is important in many situations, but it is vital when tests are being used to make important decisions that affect people's lives (e.g. recruitment and development decisions). Good psychometric tests have the advantage that their error is made explicit. In many other forms of assessment, no recognition of error is made and test scores or results are treated as absolute truths. A good example of this is exam grades or degree classes, which often contain more error than psychometric tests despite there being no acknowledgement of this error.

According to classical test theory, any test score is made up of two components: true score and error score. A person's true score is their hypothetical score on the trait being measured. For the PfS-Reasoning Tests, the true scores refer to a person's Verbal, Numerical or Abstract reasoning ability. However, scores obtained from tests also contain an error component. Error in test scores can come from three sources: the test itself, the person taking the test and the situation in which the test is being taken.

- Test error – Classical test theory assumes tests are made up from a sample of items taken from the universe of all possible items. As with any sample, this will contain a degree of error. As all people taking a test answer the same set of items, test error is systematic error, being the same for each test taker. Providing that adequate content validity has been ensured, test error is less of a concern to test users than individual or situational error.

© 2003 – 2013 Team Focus Limited

- Individual error – The individuals who take the tests are a source of random error. Factors such as how the person is feeling, their motivations and attitudes towards the testing session, and their familiarity with tests and the test format will all affect how they perform, but are not necessarily related to their actual ability. Sending out the Test Taker's Guide is one way to help limit the effect of individual error, as it ensures all test takers have a chance to become familiar with the format of the tests and know how to prepare for the test session.

- Situational error – The actual test session itself is a further source of random error. The guidelines on administration and the standardised instructions aim to make each testing session as similar as possible for all test takers. However, it is not possible to standardise the testing situation completely. The rooms used for testing, environmental conditions, time of day and interaction between the administrator and test takers will all vary between sessions. Each of these factors can influence test performance but are not related to the test taker's true ability.

Reliability statistics

In practice, the reliability of a test is typically assessed in three ways. The first of these is to look at how the test items hang together to form a coherent assessment of the construct under consideration. This 'internal consistency' is found by taking the mean of the correlation between each test item and total test score, excluding that item. Internal consistency is calculated through a formula known as Cronbach's Coefficient Alpha (or Kuder-Richardson 20 (KR20) when test items are dichotomous) and expressed as a statistic that can range from 0 to 1. The closer to 1, the more reliable the test is said to be.

The second way in which reliability is assessed is through looking at how consistent results are over time. This is done through administering the test at one point in time and then again sometime later. The scores from the two administrations are then correlated with each other to give an indication of 'test-retest' reliability. As with internal consistency, the closer the test-retest correlation coefficient is to 1, the more reliable the test is seen to be.

A further way in which reliability can be assessed is through parallel, or alternate, forms of the test. Alternate versions of the same tests can be particularly useful in applied settings, where it may be desirable to administer the same test more than once or to use a less exposed version. Typically, parallel forms are administered back-to-back and the results from the two are correlated, as when assessing test-retest reliability.

© 2003 – 2013 Team Focus Limited

Each of the statistics described above provides an index of reliability, but does not directly indicate the degree of error in a given test score. The standard error of measurement (SEM) provides a way of quantifying the error in a test score, indicating the range within which a person's true score is likely to fall. The SEM is derived from the following formula:

$$SEM = SD\sqrt{1-r}$$

where the *SD* is the standard deviation of the test in raw score units and *r* is the reliability (in this case internal consistency) of the test.

The SEM is used to create 'confidence bands' around test scores. It is known that a person's true score is likely to fall within one SEM either side of their observed score, 68% of the time. This range of scores around an observed score is known as a 'confidence band'. By multiplying the SEM by 1.28, 1.65 or 2, the confidence band can be increased to 80%, 90% or 95%. Using these values it is possible to be 80%, 90% or 95% certain that a person's true score will fall within the confidence band. In the norm tables given in **Appendix Three**, 68% and 80% confidence bands around the T-scores are given. The following sections present evidence on the internal consistency, test-retest reliability and parallel form reliability.

Internal consistency

Table 6 shows the descriptive statistics, internal consistency and SEM for each of the Reasoning Tests. A number of factors can affect the reliability and the SEM statistics. A brief discussion of the two main factors follows to allow tests users to understand the statistics given in **Table 6** more fully.

The length of a test is an important determinant of its reliability. Classical test theory assumes that any test is made up of a sample of items from the domain being assessed. As with any sample, the results from it should be more accurate as the sample becomes larger. Hence, there is a trade-off between reliability and practicality: high reliability is desirable, but if a test takes a long time to complete, very few people will choose to use it.

It is possible to construct highly reliable tests of manageable length, by developing them carefully. The rigorous development process for the PfS-Reasoning Tests is described in **Section Four**. As development was done using computer-based tests this also allowed timing data on each test item to be gathered during the trialling stage, meaning that time-efficient items were selected for the final tests. This has resulted in the timed part of the PfS-Reasoning Tests – between 10 and 15 minutes for the closed tests and 15 and 20 minutes for the open tests – being less than many similar tests of equivalent or even lower reliability.

© 2003 – 2013 Team Focus Limited

Another factor that affects reliability is the time limit allowed for the test. When tests are highly speeded, reliability estimates tend to become inflated. The item analyses indicate that the time limits allowed for each of the tests to be fairly generous, with the 'not reached' figures being similar to comparable tests. Reliability estimates are therefore unlikely to be unduly affected by the timing of the tests.

© 2003 – 2013 Team Focus Limited

Test and level		Mean	SD	Sample size	Number of items	Internal consistency	SEM
Closed Reasoning Tests							
Verbal	1	16.62	5.73	210	32	0.90	1.81
	2	16.32	5.18	303	32	0.80	2.32
	3	24.10	6.07	1322	40	0.86	2.27
	4	25.45	6.27	1131	40	0.87	2.26
Numerical	1	19.30	4.64	250	28	0.93	1.23
	2	14.95	4.74	337	28	0.84	1.90
	3	18.04	5.69	1609	36	0.87	2.05
	4	16.24	6.50	1510	36	0.89	2.16
Abstract	1	28.51	7.82	156	50	0.93	2.07
	2	20.80	8.24	242	50	0.87	2.97
	3	31.20	11.18	860	60	0.92	3.16
	4	30.35	10.41	881	60	0.91	3.12
Open Reasoning Tests							
Verbal	1	14.90	12.37	1010	44	0.92	3.50
	2	29.61	10.32	24072	60	0.91	3.10
	C*	13.75	4.70	763	24	0.84	1.88
Numerical	1	14.45	10.76	1356	40	0.92	3.04
	2	18.31	6.48	37241	48	0.85	2.51
	C	12.35	4.15	763	20	0.86	1.55
Abstract	1	39.04	13.47	515	70	0.95	3.01
	2	33.69	11.67	13.61	75	0.92	3.30
	C	16.83	5.99	763	35	0.85	2.32

* Combined Reasoning Test

Table 6: Mean, SD, sample size, number of items, internal consistency and SEM for the PfS-Reasoning Tests

© 2003 – 2013 Team Focus Limited

Test-retest reliability

Evidence of the test-retest reliability for the PfS-Reasoning Tests has been obtained from a client who requested bespoke versions of the Verbal, Numerical and Abstract tests for their selection process. These tests consist of items from the Levels 2, 3 and 4 closed tests plus other items taken from the Reasoning Test item bank. The tests taken by candidates at this organisation are computer-based and taken under supervised conditions. The organisation's policy allows candidates to re-apply after a period of time if they are initially unsuccessful, so giving a subset of applicants who have two sets of Reasoning Test data.

The sample for the test-retest analysis consisted of 169 candidates who first completed the tests April 2003 and May 2005, and completed them for the second time (retest) between July 2003 and November 2005. One hundred and thirty seven (81.1%) were male and 32 (18.9%) were female. Mean age at time of first testing was 21.4 years (SD=3.8). The mean length of time between first taking the tests and retesting was 38.7 weeks (SD=25.3 weeks), with a range from 2 days to 121 weeks. For the majority of candidates retesting occurred between 10 and 40 weeks after first taking the tests.

	First time (n=169)		Retest (n=169)				
	Mean	SD	Mean	SD	Difference	Test-retest correlation	Number of test items
Verbal	26.7	5.2	29.0	5.0	2.3	0.73	40
Numerical	18.1	3.8	19.6	3.6	1.5	0.71	36
Abstract	42.3	9.5	49.6	9.9	7.3	0.67	70

Table 7: Mean and SD for first time and retest candidates, and test-retest reliabilities for bespoke versions of the PfS-Reasoning Tests

Test test-retest correlation coefficients showed that each of the tests had adequate reliability, particularly considering the extended time between testing for many in the sample – almost half a year on average. It should also be noted that the test-retest correlations in **Table 7** are likely to underestimate the true correlations, as these have not been corrected for measurement error.

The data in **Table 7** also gives an indication in likely change in scores on retest. Mean scores on all three tests increased on retesting, although the standard deviation remained relatively constant. An indication of the magnitude of score change can be obtained by looking at the absolute change in mean test score as a proportion of the test's standard deviation (taken from all 5294 candidates from this organisation). From these calculations values of 0.43, 0.38 and 0.71 for the Verbal, Numerical and Abstract tests respectively were obtained, showing modest increases in mean Verbal and Numerical scores and a slightly larger increase in Abstract scores of just under three quarters of a standard deviation.

© 2003 – 2013 Team Focus Limited

One possibility for the modest changes in score is the extended time period before retesting for some candidates (over 2 years in a few cases). Correlations between the difference between first time and retest scores, however, gave no indication that these scores were associated with time between the two test sessions.

Standard error of difference

Given that all test scores contain a degree of error, one important question which test users often ask is "Are the scores of two people really different?". If test scores were free from error, any difference in observed scores would reflect a real difference in the ability being assessed. However, because of error, if two scores are close to each other, there is a chance that they could be reversed if the test takers took the tests again. In other words, the person who obtained the higher score may not continue to obtain the higher score the second time around.

The likelihood of the difference between two test scores reflecting a real difference in the construct being assessed, can be determined with a statistic known as the standard error of difference (SED). The SED indicates how far two test scores need to be apart before the difference can be seen as meaningful. The formula for the SED is:

$$SED = \sqrt{SEM_1^2 + SEM_2^2}$$

where SEM_1 is the standard error of measurement for the first test and SEM_2 is the standard error of measurement for the second test. Using this formula one person's scores on different tests can be compared. This can be particularly useful when tests are being used to identify an individual's relative strengths and weaknesses, possibly for development purposes. In selection situations it is more common to compare different people's scores on the same test. In this situation, SEM_1 and SEM_2 have the same value, meaning that the formula can be simplified to:

$$SED = 1.414 \times SEM$$

As with the SEM, if two scores differ by one SED or more, the higher scorer is likely to remain on top 68% of the time – about two times out of three. Alternatively, this situation can be expressed as being 68% certain that there is a real difference between the scores. By multiplying the SED by 1.28 or 2.0, the level of certainty can be increased to 80% or 95% that two people's scores really are different. The SED in raw scores and T-scores for each of the Reasoning Tests is shown in **Table 9** below.

© 2003 – 2013 Team Focus Limited

Closed tests							
		68% SED		80% SED		95% SED	
		Raw score	T-score	Raw score	T-score	Raw score	T-score
Verbal	1	2.56	4.47	3.28	5.72	5.12	8.94
	2	3.28	6.32	4.19	8.09	6.55	12.65
	3	3.21	5.29	4.11	6.77	6.42	10.58
	4	3.20	5.10	4.09	6.53	6.39	10.20
Numerical	1	1.74	3.74	2.22	4.79	3.47	7.48
	2	2.68	5.66	3.43	7.24	5.36	11.31
	3	2.90	5.10	3.71	6.53	5.80	10.20
	4	3.05	4.69	3.90	6.00	6.10	9.38
Abstract	1	2.93	3.74	3.74	4.79	5.85	7.48
	2	4.20	5.10	5.38	6.53	8.40	10.20
	3	4.47	4.00	5.72	5.12	8.94	8.00
	4	4.42	4.24	5.65	5.43	8.83	8.48
Open tests							
Verbal	1	4.95	4.00	6.33	5.12	9.89	8.00
	2	4.38	4.24	5.60	5.43	8.76	8.48
	C*	2.66	5.66	3.40	7.24	5.32	11.31
Numerical	1	4.30	4.00	5.51	5.12	8.61	8.00
	2	3.55	5.48	4.54	7.01	7.10	10.95
	C	2.20	5.29	2.81	6.77	4.39	10.58
Abstract	1	4.26	3.16	5.45	4.05	8.52	6.32
	2	4.67	4.00	5.97	5.12	9.33	8.00
	C	3.28	5.48	4.20	7.01	6.56	10.95

Table 8: SED for the PfS-Reasoning Tests at 68%, 80% and 95% confidence levels

In order to use **Table 7**, first identify the confidence level required (68%, 80% or 95%) and whether raw scores or T-scores are being used. Find the appropriate column using the first two rows of **Table 7**. Then find the appropriate test in the left-hand column and follow the row across until it intersects with the column to obtain the SED. Test scores need to differ by at least the SED before the difference can be said to be real. For example, to be 80% certain that raw scores from Numerical test Level 2 closed test reflect a real difference in numerical reasoning ability, the difference between raw scores has to be at least 3.44 points. The values in **Table 7** are given as decimals whereas test scores will typically be whole numbers. If users wish to work with whole numbers for simplicity, SEDs should always be rounded up and never down, as rounding down will reduce the confidence that can be placed in the SED. Rounding up will effectively make no difference, as test scores are whole numbers.

© 2003 – 2013 Team Focus Limited

Bias

When used appropriately, psychometric tests have the potential to offer objective, unbiased assessments of ability, aptitude or personal characteristics. Bias in testing occurs because tests have been poorly constructed or because they are used inappropriately. An overview of how to select tests appropriately has been given in **Section Two**. Ensuring that all people are tested under the same conditions by following the standardised administration procedure further reduces the possibility of bias.

More fundamental than appropriate test use is test construction; if a test is inherently biased, the results it gives will always be biased regardless of whether it is being used and administered appropriately. Test bias can arise when the test measures the construct it purports to, but also another, unrelated construct. If the level of this unrelated construct varies between different groups, then the overall results from the test may be biased. For example, a numerical test may contain a large verbal component. If verbal ability differs between two groups (say, people with and without English as their first language) scores may favour one group over another, even if the assessment of numerical ability within the test is fair to both groups.

The initial development of the PfS-Reasoning Tests involved the definition of the areas to be assessed and identification of appropriate test formats (see **Section Four**). From this definition the test specifications were developed, including the descriptions of suitable item content for each Reasoning Test. Bias was therefore minimised by ensuring that the tests did not assess constructs other than the core Verbal, Numerical or Abstract reasoning abilities. Test items were also reviewed for possible bias and subjected to bias analyses during the trailing stage.

Bias can be assessed in two ways: through an examination of overall test scores or the difficulty of individual test items. To assess whether differences in total test scores indicate bias or reflect real differences in the constructs being assessed, it is necessary to find a marker against which test scores can be assessed. As pure markers for constructs such as reasoning abilities are very difficult to identify, the item-level approach to bias was used in the development of the PfS-Reasoning Tests.

The item-level bias analyses conducted during the development of the PfS-Reasoning Tests used a technique known as differential item functioning (*dif*). *Dif* analyses identify whether individual test items are found to be disproportionately easy or hard by different groups of test takers, once their overall score on the test has been allowed for. In other words, if two groups of test takers (say, males and females) obtain very similar overall test scores, the chances of them answering each item correctly should be approximately the same. If one group has a much higher chance of answering an item correctly, the item may be biased.

© 2003 – 2013 Team Focus Limited

Dif analyses require quite large samples for the results to be robust. Analyses were conducted during the initial stages of development for males and females and for test takers who described their ethnic background as being 'white' and those from other ethnic backgrounds ('non-whites'). More detailed *dif* analysis of the specific ethnic groups was not possible during the initial stages of development due to the large samples required to do this reliably. Few items were seen to show significant *dif*, suggesting that the pre-trialling reviews and screening of items prior to constructing the final versions of the tests had successfully identified problematic items.

Mean test scores were also examined for males and females and 'whites' and 'non-whites'. The results of these are shown in **Tables ~~0~~9 and 10** below. As can be seen from **Table ~~0~~9**, significant score differences were observed between males and females on a number of the PfS-Reasoning Tests. With relatively large sample sizes, however, even very small differences between groups can reach statistical significance. Because of this it is more appropriate to examine differences in terms of 'effect sizes', which look at the difference between groups as a proportion of the pooled standard deviation (taken from **Table 6**). Effect sizes are shown in the last columns of **Tables ~~0~~9 and 10**.

Guidelines for interpreting effect sizes describe values less than 0.2 as indicating 'small' differences between groups, those between 0.2 and 0.5 as 'medium' and those above 0.5 as 'large' (Cohen, 1988). All differences between males and females fall into the 'small' or 'medium' effect sizes, though there is no consistent pattern of differences between the different test types. This suggests that many of the observed differences may be due to the characteristics of specific samples.

Comparisons of the mean test scores of 'whites' and 'non-whites' revealed a number of statistically significant differences and a number of cases where the effect sizes associated with these differences were of either a medium (10 out of 21 comparisons) or large (5 out of 21 comparisons) magnitude. The remaining 6 comparisons were of a small magnitude. In all cases the differences were seen to favour the 'white' group over the 'non-white' group. These findings reflect the well-established evidence that people from ethnic minority groups, on average, tend to achieve lower scores on ability tests (e.g. College Board, 2003).

Simple comparisons such as 'whites' and 'non-whites' can mask more complex patterns of performance seen between more precisely defined groups, but to make such comparisons usually requires large numbers of test takers so that all groups are adequately represented. Such comparisons were possible with the open level 2 Reasoning Tests, which have been used extensively in universities and for which large samples of test takers from more precisely specified ethnic groups were available. The results of this analysis can be seen in **Table 11**, which shows the mean test scores according to the 16 ethnic groups defined for the 2001 census of England and Wales.

© 2003 – 2013 Team Focus Limited

		Males			Females			Difference	Effect size
		Mean	SD	Sample size	Mean	SD	Sample size		
Closed Reasoning Tests									
Verbal	1	14.99	5.75	78	17.58	5.51	132	2.59**	0.45
	2	16.20	5.06	158	16.44	5.32	145	0.24	0.05
	3	24.34	6.28	614	23.90	5.88	708	0.44	0.07
	4	25.68	6.46	580	25.22	6.07	551	0.46	0.07
Numerical	1	18.87	5.01	133	19.79	4.16	117	0.92	0.20
	2	15.20	4.57	201	14.57	4.97	136	0.63	0.13
	3	18.69	5.93	864	17.29	5.30	745	1.40***	0.25
	4	16.87	6.71	883	15.35	6.10	627	1.52***	0.23
Abstract	1	27.55	8.15	75	29.41	7.43	81	1.86	0.24
	2	18.82	7.96	125	22.91	8.02	117	4.09***	0.50
	3	30.81	11.66	446	31.61	10.65	414	0.80	0.07
	4	30.64	10.56	503	29.98	10.21	378	0.66	0.06
Open Reasoning Tests									
Verbal	1	17.03	12.21	434	13.29	12.25	576	3.74***	0.30
	2	29.60	10.27	12813	29.62	10.39	11259	0.02	0.01
	C[#]	13.30	4.81	396	14.24	4.45	367	1.06**	0.23
Numerical	1	13.75	11.20	716	15.42	10.18	640	1.67**	0.16
	2	18.67	6.64	20966	17.85	6.23	16275	0.82***	0.13
	C[#]	12.18	4.43	396	12.54	3.83	367	0.36	0.09
Abstract	1	37.85	13.45	301	40.73	13.34	214	2.88*	0.21
	2	33.04	11.99	7349	34.54	11.20	5682	1.50***	0.13
	C[#]	15.66	5.57	396	18.09	6.17	367	2.43***	0.41

[#]Combined Reasoning Test
*p<0.05; **p<0.01; ***p<0.001

Table 09: Mean raw scores and standard deviations for males and females on the PfS-Reasoning Tests

© 2003 – 2013 Team Focus Limited

		'whites'			'non-whites'			Difference	Effect size
		Mean	SD	Sample size	Mean	SD	Sample size		
Closed Reasoning Tests									
Verbal	1	19.09	6.49	708	16.81	6.27	79	2.28**	0.40
	2	18.80	5.31	233	14.22	6.08	27	4.58***	0.88
	3	25.43	5.60	817	22.25	6.17	228	3.18***	0.52
	4	26.46	5.77	772	23.47	6.94	230	2.99***	0.48
Numerical	1	20.17	5.69	721	19.56	5.42	94	0.61	0.06
	2	15.62	5.67	273	13.72	4.44	36	1.90	0.26
	3	18.79	5.83	989	16.74	5.54	324	2.05***	0.44
	4	16.25	6.35	1028	15.57	6.76	306	0.68	0.10
Abstract	1	28.12	9.80	547	27.97	8.20	61	0.15	0.02
	2	24.18	8.06	155	19.65	5.51	20	4.53*	0.55
	3	31.80	11.25	531	28.96	10.95	102	2.84***	0.25
	4	30.60	11.00	641	26.78	10.02	141	3.82***	0.37
Open Reasoning Tests									
Verbal	1	22.19	11.98	388	10.05	10.21	564	12.14***	0.98
	2	32.22	9.46	14389	25.54	10.08	8622	6.68***	0.65
	C[#]	13.94	4.65	680	12.05	4.91	76	1.89***	0.42
Numerical	1	16.27	10.65	675	12.82	10.60	681	3.45***	0.32
	2	18.37	6.06	22349	18.20	7.02	13372	0.17*	0.03
	C[#]	12.47	4.10	680	11.39	4.51	76	1.08*	0.26
Abstract	1	41.05	12.54	319	35.67	13.78	180	5.38***	0.40
	2	34.47	11.03	8124	32.26	12.34	4346	2.21***	0.19
	C[#]	16.98	5.84	680	15.87	7.05	76	1.11	0.19

[#]Combined Reasoning Test
*p<0.05; **p<0.01; ***p<0.001

Table 10: Mean raw scores and standard deviations for 'whites' and 'non-whites' on the PfS-Reasoning Tests

© 2003 – 2013 Team Focus Limited

	Verbal				Numerical				Abstract			
	N	Mean	SD	Effect size	N	Mean	SD	Effect size	N	Mean	SD	Effect size
British	10499	32.83	9.39		15456	18.46	5.96		5412	34.88	10.76	
Irish	1213	31.84	9.01	0.10	2029	17.78	5.44	0.10	641	33.30	10.80	0.14
Any other White background	2677	30.00	9.60	0.27	4864	18.34	6.57	0.02	2071	33.78	11.72	0.09
White and Black Caribbean	144	19.83	11.84	1.26	199	13.54	7.07	0.76	125	27.59	11.21	0.62
White and Black African	62	29.13	11.12	0.36	108	15.59	5.67	0.44	61	29.66	13.42	0.45
White and Asian	247	28.38	12.13	0.43	321	18.80	6.48	0.05	120	36.78	13.00	0.16
Any other mixed background	246	29.99	10.39	0.28	391	17.96	6.30	0.08	150	34.10	10.64	0.07
Indian	2126	26.82	9.93	0.58	3715	17.25	6.67	0.19	1132	30.14	12.19	0.41
Pakistani	406	27.44	8.88	0.52	643	16.68	6.18	0.27	168	29.97	10.70	0.42
Bangladeshi	184	25.86	9.47	0.68	262	17.29	6.60	0.18	76	33.67	10.73	0.10
Any other Asian background	760	25.79	9.89	0.68	1154	17.75	6.99	0.11	417	31.38	10.45	0.30
Caribbean	198	24.40	9.14	0.82	308	14.80	6.31	0.56	98	28.48	10.49	0.55
African	656	24.12	9.17	0.84	1090	15.84	6.00	0.40	360	27.25	9.76	0.65
Any other Black background	65	25.58	7.52	0.70	90	15.40	5.98	0.47	31	26.84	9.18	0.69
Chinese	3289	24.27	9.97	0.83	4718	20.38	7.16	0.30	1463	35.59	12.96	0.06
Any other	239	27.20	10.65	0.55	373	18.07	7.12	0.06	145	35.17	12.49	0.02

Table 11: Mean test scores and effect sizes for different ethnic groups based on the open Level 2 PfS-Reasoning Tests Reasoning Tests

The detailed analysis of test scores obtained by different ethnic groups shown in **Table 11**, indicates the means and SDs for each group on the Verbal, Numerical and Abstract level 2 open tests. The effect size for each group is also shown, and indicates the extent to which the mean for each group differs from the 'British' mean.

© 2003 – 2013 Team Focus Limited

The ‘British’ group obtained the highest mean score on the Verbal test, with the ‘Chinese’ group obtaining the highest mean score on the Numerical and Abstract tests. In terms of lower scoring groups, ‘White and Black Caribbean’, ‘Caribbean’, ‘African’ and ‘Any other Black background’ consistently showed some of the largest effect sizes.

Additional work in this area has been conducted with a professional body recruiting members to a practising panel. This work was conducted in 2011-2012 with 217 postgraduate calibre applicants completing closed Level 3 tests.

The distribution of those 'appointed' to 'not appointed' is given in Table 12. However of more interest is the examination of the possible ethnic group differences with regard to the three ability tests (Table 13)

	White		Non-white or mixed		Total no
	No.	%	No.	%	
Not appointed	118	59.90%	13	65.00%	131
Appointed	79	40.10%	7	35.00%	86
Totals	197	90.78%	20	9.22%	217

Table 12: Cross tabulation between appointment decision and aggregated ethnic group categories

It can be seen that 40.1% of candidates falling in the White group were successful compared to 35% of those falling in 'None-white or mixed' group. A Chi-square computed on the data in Table 3 showed that, although the proportions of successful candidates differ between the white and none-white / mixed groups, they do not differ significantly (Chi-square= 0.1975, df=1, p=0.65674).

In order to examine possible ethnic group differences on the three psychometric test scores, a multivariate analysis of variance (MANOVA) was carried out, using the raw scores on the three tests (numerical, verbal and abstract reasoning) as the dependent variables and ethnic group as the independent variable. The mean raw scores on the three tests are presented in Table 13.

© 2003 – 2013 Team Focus Limited

Ethnic	abstract	numerical	verbal	N
White - English	34.75	21.55	31.80	55
White - British	35.42	21.72	32.40	108
White - Irish	28.14	22.57	31.43	7
White - Welsh	37.60	22.00	32.00	5
White - Scottish	29.33	21.22	32.00	9
White - Other	34.18	20.55	31.82	11
Asian or Asian British - Indian	33.20	20.70	31.60	10
Asian or Asian British - Pakistani	32.00	22.50	35.00	2
Asian or Asian British - Bangladeshi	37.50	23.00	31.00	2
Asian or Asian British - Other	14.00	19.00	32.00	1
Chinese	42.00	23.00	29.00	1
Black or Black British - Caribbean	44.00	34.00	35.00	1
Mixed - Other	7.00	25.00	29.00	1
Mixed - White and Asian	33.50	18.50	31.50	2

Table 13: Raw score means on the three reasoning tests for each ethnic group

Some of the sub-groups contained very few individual. However the overall main effect of ethnic group was found to be non-significant (Wilk's Lambda=.80805, p=.27567) indicating that there was no overall difference in performance on the three tests between individuals from different ethnic groups.

Also with the recent introduction of legislation on age, there has been particular interest in how performance on tests of mental ability such as the PfS-Reasoning Tests is influenced by age. The links between test scores and age are seen in **Table 14**. The strongest links between age and test performance are seen the closed level 1 tests and the Combined test. These positive correlations indicate that test scores increase as does respondents' age, and most likely reflects progress and development through the education system as these tests were taken by people as young as 14. The largest of these associations, seen for the closed level 1 Verbal test, indicates that age accounts for just over 8% of the variance in test scores. Some evidence of a negative association with age was also seen amongst samples with slightly older respondents, though there was no evidence of a highly significant fall-off in performance. Taken across all test levels, age accounted for less that 2% in performance on average.

© 2003 – 2013 Team Focus Limited

Test version	Verbal	Numerical	Abstract
Closed Tests			
1	0.29 (n=803)	0.23 (n=831)	0.04 (n=598)
2	0.03 (n=549)	0.15 (n=595)	-0.15 (n=429)
3	-0.06 (n=1340)	-0.10 (n=1624)	-0.12 (n=878)
4	0.13 (n=1137)	0.05 (n=1552)	0.02 (n=889)
Open tests			
1	0.22 (n=1010)	-0.10 (n=1343)	-0.13 (n=512)
2	-0.02 (n=23999)	-0.13 (n=37134)	-0.13 (n=12981)
Combined	0.22 (n=755)	0.18 (n=755)	0.14 (n=755)

Table 14: Associations between raw PfS-Reasoning Tests and respondents age

© 2003 – 2013 Team Focus Limited

A commentary on interpreting bias data

When variations in test scores are seen between different groups, whether those groups are defined on the basis of sex, ethnicity, age or any other factor, an immediate possibility is that the tests are biased. That is, in some way, the items within the test or the whole testing process itself are easier for some groups than others, resulting in differential performance. Such differential performance is not in itself a bad thing, but becomes an issue if it can be shown that the differences arise due to test performance being affected by extraneous factors unrelated to the construct being assessed – in the current case, Verbal, Numerical or Abstract reasoning ability.

With a focus on the analysis of ethnic groups performance, as it is here that the largest mean differences were seen, the purpose of this section is to explore the possible reasons for this differential test performance. It is recognised, and largely accepted, that variations in test performance will be seen between different ethnic groups. These differences remain despite the best efforts of test developers to make tests fair and accessible through careful test design, item writing and review, trialling and item-level statistical analyses. Comparisons between the current tests and other aptitude tests will therefore indicate the extent to which the PfS-Reasoning Tests can be considered as functioning within 'accepted' parameters.

The Scholastic Assessment Test (SAT) used as part of college selection in America and other countries for over 2 million students each year, is probably the most researched and well-developed aptitude testing programme, and so provides an appropriate benchmark for the examination of ethnicity. ETS, who develop the SAT, have also been influential in shaping modern thinking on test bias and how to identify it.

In terms of effect sizes, a difference of 0.98 (almost 1 SD) is seen between 'White' and 'African American' candidates on the verbal part of the SAT and a difference of 1.08 seen on the math part, with 'African American' candidates scoring lower in both cases. When compared to the 'Asian American' group, 'White' candidates score 0.21 higher on verbal, but 0.41 lower on math (CollegeBoard, 2003). These figures indicate that substantial differences between the mean scores of different ethnic groups remain, despite the best efforts of test developers. They are also in line with the findings from **Table 11**, where the 'White and Black Caribbean', 'Caribbean', 'African' and 'Any other Black background' groups showed some of the largest effect sizes when compared with 'Whites' and obtained some of the lowest mean test scores.

It is currently unclear why these differences are seen, although there are a number of possibilities (see for example Freedle, 2003 and Neisser, Boodoo, Bouchard, Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg and Urbina, 1996 for a discussion). First, any differences may reflect true differences in the capability of candidates. As there is no 'gold standard' against which aptitudes can be measured, it is very difficult to establish any individual's or group's true level of specific abilities. The only robust way of checking a test for bias, and so determining whether test score differences

© 2003 – 2013 Team Focus Limited

reflect differences in the ability to perform a job, is through comprehensive validity studies. Second, the possibility of 'differential sampling' needs to be considered. The effect of any background factors on test scores could be due to groups being made up from people of different ability levels. To determine whether differential sampling is affecting the observed scores it would be necessary to collect additional background information on candidates, particularly educational qualifications and proficiency in English. Third, differences could be due to variations in familiarity with reasoning tests. Finally, a number of cultural and sociological arguments have been proposed to explain differential test performance (see Neisser *et al* 1996 for a summary). Many of these theories focus on the meaning and experience of testing to people from different cultures, recognising that the whole testing movement has its roots in a 'white, middle-class' philosophy.

To summarise, differences in the mean test scores of different groups do not prove that a test is based. The ethnic differences observed in the PfS-Reasoning Tests are also seen in other widely used tests and remain despite intensive efforts to make tests 'fair'. Ensuring the fairness of the tests is an ongoing project combining test research, support to candidates and the need to validate the tests against meaningful and reliable job-related criteria. Further analyses will therefore be conducted and reported in subsequent versions of this User's Guide when sufficient data is available.

Validity

Validity is the most important consideration when using any test or assessment. If a test is valid, it will produce meaningful results and will contribute significantly to the decision-making process, either predicting subsequent job or training performance or correctly identifying development needs.

Many different forms of validity have been identified, but it is most accurately viewed as a unified concept (Messick, 1995), with different forms of validity contributing to an overall judgement. Further, it can never be asserted that a test is globally valid or not, as validity relates to the use of a test in specific situations – its 'fitness for purpose'. Four main types of validity are discussed here: face, content, construct and criterion validity.

Face validity

A test has face validity when it looks as though it is measuring what it claims to measure. Although not always considered to be a genuine source of validity, if test takers can clearly see links between the skills being measured and a certain job, they are likely to be motivated to complete the test to the best of their abilities. There may be lower motivation to perform well if the reasons for completing the test are unclear. Further, the selection process is seen as a form of social interaction, during which applicants will form impressions of an organisation (Anderson and Cunningham-Snell, 2000). The use of tests with low face validity may have a negative impact on this emerging impression. For these reasons, face validity is important when using assessments in occupational settings.

© 2003 – 2013 Team Focus Limited

Evidence for the face validity of the PfS-Reasoning Tests was collected during the trialling stage by observing test sessions and obtaining feedback from test takers. Users found the tests easy to use and the content to be acceptable. Further, the feedback reports designed for test takers were also seen to be accessible, informative and to provide useful points for consideration. However, some users found the reports to be quite long and suggested that reports simply containing the test results would have been more useful for their purposes. To address this need, summary reports for both administrators and test takers were created (see **Appendix Two** for a sample). Although feedback indicated that the tests had good face validity, this has to be supported by other forms of validity or a test may be accused of superficiality.

Content validity

If the items in a test provide adequate coverage of the area being assessed, and do not relate to areas outside the sphere of the test, then the test is said to have content validity. For the Verbal, Numerical and Abstract Reasoning Tests, the process of ensuring content validity started by developing the test specifications, detailing the content of each test. The review of test items by the test development team and external experts further contributed to content validity, ensuring that items met the test specifications and making necessary changes where they did not. The final stage in this process was the compilation of the tests themselves, where the content of each separate test was carefully checked to make sure it was adequate.
The development process has resulted in tests which fulfil the specifications set out in **Section Four**. Ultimately, however, potential test users should review the tests themselves, to ensure that test content sufficiently matches their needs.

Construct validity

Construct validity refers to what a test actually measures. In the case of the PfS-Reasoning Tests, the constructs are Verbal, Numerical and Abstract reasoning ability. Evidence for construct validity comes from the examination of how scores on each of the tests relate to each other and to established assessments that measure related constructs.

The correlations between the three PfS-Reasoning Tests at each level are shown in **Table 15**. A number of observations can be made from this data. Firstly, the correlations show that each of the Reasoning Tests is assessing a quite distinct area of ability. The highest correlation is between the Numerical and Verbal parts of the Combined Test, showing that the two share around 42% of common variance (i.e. performance on one test will account for no more than 42% of the performance on another). Among the closed tests the degree of association is generally far less, with the mean correlation indicating that just under 20% of common variance is shared between tests.

Secondly, there is a decrease in the mean correlations between the higher levels of the closed tests. The mean correlations are 0.56, 0.45, 0.45 and 0.32 for levels 1 to

© 2003 – 2013 Team Focus Limited

4 respectively. It is known that as people get older and specialise in their areas of study abilities tend to become more defined, meaning that the correlations between assessments of different abilities are reduced. The pattern of relationships found with the PfS-Reasoning Tests supports this differentiation of abilities. This observation is further supported by the data from the Combined Reasoning Test, where the majority of test takers were in the last two years of compulsory education. The level of correlation in this test is also likely to be influenced by the fact that the three sub-sections are relatively short and taken immediately after each other, so potentially reducing some of the sources of error in test scores.

Together these findings show a meaningful pattern of relationships within the three PfS-Reasoning Tests, indicating that they assess quite distinct areas of reasoning ability and so supporting the validity of the constructs defined for the PfS-Reasoning Tests.

© 2003 – 2013 Team Focus Limited

Closed Reasoning Tests

	Verbal 1	Numerical 1
Numerical 1	0.60 (189)*	-----
Abstract 1	0.53 (131)	0.54 (120)

	Verbal 2	Numerical 2
Numerical 2	0.53 (263)	-----
Abstract 2	0.41 (237)	0.39 (241)

	Verbal 3	Numerical 3
Numerical 3	0.48 (1659)	-----
Abstract 3	0.48 (1304)	0.40 (1218)

	Verbal 4	Numerical 4
Numerical 4	0.28 (1240)	-----
Abstract 4	0.35 (805)	0.33 (813)

Open Reasoning Tests

	Verbal 1	Numerical 1
Numerical 1	0.37 (1288)	-----
Abstract 1	0.38 (106)	0.44 (118)

	Verbal 2	Numerical 2
Numerical 2	0.38 (5820)	-----
Abstract 2	0.47 (3565)	0.35 (3731)

	Combined - Verbal	Combined - Numerical
Combined - Numerical	0.65 (880)*	-----
Combined - Abstract	0.56 (880)	0.62 (880)

* figures in parentheses indicate number of test takers

Table 15: Intercorrelations of the PfS-Reasoning Tests

A more recent data set, based on data gathered over the course of 2012 from respondents who had completed more than one version of each test, produces a similar pattern of results for the closed tests. With Numerical and Verbal, Level 1, 2 and 3 tests, yielding intercorrelations of 0.61 (Level 1, N-167), 0.60 (Level 2, N=156) and 0.55 (Level 3, N=195); Numerical and Abstract, Level 2 and 3 tests, intercorrelations of 0.50 (Level 2, N=127) and 0.53 (Level 3, N=125); and Verbal and Abstract, Level 2 and 3 tests, intercorrelations of 0.47 (Level 2, N=131) and 0.51 (Level-3, N=158).

© 2003 – 2013 Team Focus Limited

Further evidence for the criterion validity of the PfS-Reasoning Tests comes from a number of studies that have explored the association between them and other assessments of capability. These studies are summarised below.

Association between level 4 closed tests and the Graduate Management Admission Test (GMAT), which is used by graduate management schools in many countries as part of their admission process, was examined in a sample of postgraduate students at a business school based in London during 2004. The sample consisted of approximately 56% males and 44% females, with a mean age of 26.39 years (SD=4.47). A significant proportion of the students in this sample came from outside of the UK, though exact data on this was not available.

As shown in **Table 16**, the strongest association was seem between GMAT and the PfS Verbal test. This would be expected as the GMAT contains two sections of verbal material and one of numerical. There is no equivalent in the GMAT to the PfS Abstract test, as reflected in the lower association of PfS Abstract with the GMAT scores. It should also be noted that respondents in this sample were asked to recall their GMAT scores from memory and that the time between taking the two assessments could have been around a year for some students, both of which could have affected the resulting correlations.

	Correlations with GMAT and sample size
Verbal 4	0.34 (n=74)
Numerical 4	0.23 (n=97)
Abstract 4	0.15 (n=54)

Table 16: Associations between PfS-Reasoning Tests and the GMAT

The Graduate and Managerial Assessment (GMA; Blinkhorn, 1985) is an established and widely used test, consisting of high-level verbal, numerical and abstract tests. As with the PfS Abstract Reasoning Tests, the GMA Abstract test is based on Bongard problems (see page 32), and this study explored the association between levels 3 and 4 of the closed PfS Abstract Reasoning Tests and the GMA Abstract Test form A (GMA-A). Data was collected during the first quarter of 2007 from two groups of Year 12 students, one at a boys-only comprehensive school and another at a girls-only independent school. There were 78 participants from the boys school with a mean age of 16.7 years (SD=0.7) and 48 from the girls school with a mean age of 16.4 (SD=0.5). The order of test completion was counterbalanced.

© 2003 – 2013 Team Focus Limited

The correlations between the scores from the three tests are shown in **Table 17**, with the first figure showing the raw correlation and the second in brackets the correlation corrected for the reliability of the two tests in question. All uncorrected correlations between PfS tests and GMA are 0.64 or greater, and when corrected for reliability are 0.71 or greater. These figures indicate a good degree of association between the assessments and all exceed the 0.70 threshold, typically recognised the point at which tests can be considered to be alternate forms of each other.

	PfS 3	PfS 4	GMA lenient
PfS 4	0.73 (0.80)		
GMA lenient	0.71 (0.78)	0.64 (0.71)	
GMA harsh*	0.68 (0.86)	0.69 (0.79)	0.80 (0.89)

* Harsh scoring on the GMA-A awards one mark only if all of the five test shapes in a group have been answered correctly.

Table 17: Associations between PfS Abstract Tests and GMA Abstract form A

The association between the PfS Verbal and Numerical level 1 open tests and SHL's VMG3 (verbal reasoning) and NMG3 (numerical reasoning) was examined in a sample of employees at a UK emergency services organisation. Forty-four employees completed both verbal tests (mean age 42.32, SD=5.02) and seventy the numerical tests (mean age 42.26, SD=5.31). All test takers were male.

The employees completed the PfS tests as preparation for an internal development programme and the SHL tests subsequently as part of the programme. The correlations between the verbal tests were 0.43 (0.51 corrected for reliability) and between the numerical tests were 0.26 (0.29 corrected for reliability).

Further data relating to construct validity was obtained from versions of the PfS-Reasoning Tests constructed with items from each of the four closed test levels. These tests were developed for a client who needed to assess people across a wide range of ages and ability levels. The correlations between the versions of the PfS-Reasoning Tests and the client's existing assessments were examined for evidence of construct validity. As the reliability of the client's existing assessments was poor the correlations in **Table 18** have all been corrected for reliability.

© 2003 – 2013 Team Focus Limited

The data reported in **Table 18** was collected in 2003 from 254 candidates to the client's organisation. The mean age of the candidates was 22.2 years (SD=3.0), and 240 (94.5%) were male and 14 (5.5%) were female.

	Correlations with existing reasoning tests and sample size
Verbal reasoning correlations	0.48 (121)
Numerical reasoning correlations	0.65 (115)
Abstract reasoning correlations	0.36 (122)

Table 18: Intercorrelations between the Verbal, Numerical and Abstract Reasoning Tests and existing reasoning tests

These correlations provide further construct validity for the PfS-Reasoning Tests, particularly the Numerical Reasoning Test. The correlation for the Abstract tests is likely to have been somewhat lower due to the format of the two tests being quite different (the comparison test consisted of a number of simple, speeded processing tasks, but was claimed to assess 'innate ability'). The association between the Verbal Reasoning Tests indicates a moderate overlap. As with the Abstract, subtle differences in the abilities required for the two Verbal Reasoning Tests is likely to explain why a higher degree of association was not observed.

In terms of the relationship between the reasoning tests and other abilities, the Level 2 reasoning tests were compared with the Team Focus Memory and Attention Test (MAT). The MAT measures the ability to quickly memorise and retain information in order to apply rules or procedures to shapes or other target stimuli, in a timely and accurate manner. It is a multi-faceted online test that generates a profile of performance as individuals respond to increasingly complex instructions and screens of information.

Measures obtained from the MAT include:

- Memory: a measure of how many times a person checks the instructions. A high score (which results from the respondent checking the instructions relatively infrequently) indicates good memory (i.e. less reliance on the instructions).
- Accuracy: the total number of correct shapes that have been clicked.
- Decision Making: a measure of the number of items answered correctly per minute. High scores show people who are both fast and accurate.

The results for a sample of 208 Year-11 students, assessed during 2012, are presented below:

© 2003 – 2013 Team Focus Limited

	Memory	Accuracy	Decision Making
Numerical	0.134	0.238*	0.253*
Verbal	0.173*	0.416*	0.205*
Abstract	0.141*	0.366*	0.171*

* $p<0.05$

Table 19: Correlations between Numerical, Verbal and Abstract Reasoning Tests and Memory and Attention Test

There are significant relationships between the three Level 2 tests and aspects of memory and attention. In particular they confirm that memory is an 'enabling' ability, i.e. that being able to retain information in working memory is an important mediator of test performance, across a range of different tests.

Association between psychometric assessments of abilities and measures of working memory capacity further support argument that working memory is a vital element of both fluid and crystallised abilities. Kyllonen and Christal (1990) found that in four separate studies the association between psychometric tests of ability and tests of Baddeley's (1974) working memory model, correlations as high as 0.8.

Criterion validity

A range of criterion data is available for the PfS- Reasoning Tests, showing how test scores are associated with three stages of educational attainment: GCSE grades, UCAS points and degree grades.

The association between the Verbal, Numerical and Abstract open level 1 tests and GCSE results are shown in **Table 20**. Test results were collected from students who had finished compulsory education and were in their first year of further academic study (predominantly A and AS levels). The mean ages for the samples completing each test ranged from 16.73 years (SD=0.46) to 16.86 years (SD=0.61), and approximately 70% of respondents in each sample were male. GCSE results were collected for the three compulsory areas of study: English, maths and science.

Table 20 shows moderate and quite consistent associations between ability assessed by the PfS-Reasoning Tests and academic attainment at the age of 15 or 16, with a mean correlation across the three tests of 0.41. As this data was collected from students who had gone on to further study, the correlations may underestimate the true association due to the restricted range of GCSE grades.

© 2003 – 2013 Team Focus Limited

	GCSE English grade	GCSE maths grade	GCSE science grade
Verbal (n=48)	0.17	0.44	0.41
Numerical (n=64)	0.48	0.53	0.20
Abstract (n=66)	0.47	0.66	0.36

Table 20: Associations between GCSE English, maths and science grades and PfS-Reasoning Tests

The association between the Verbal, Numerical and Abstract open level 2 Tests, UCAS points and degree class is shown in **Table 20**. Overall, test scores showed only a modest association with UCAS points and very little association with degree class. It should be noted, however, that UCAS points were collected retrospectively from test takers, sometimes a number of years after the examinations contributing to UCAS points had been taken. Degree class showed considerable restriction in range, with between 55 and 60% of respondents indicating their degree class as being '2i'. The comparability of degrees from different institutions and in different subjects is also highly questionable, and are both further factors likely to mask any true associations.

	Sample	UCAS points	Sample	Degree class
Verbal 2	Age 22.68 (5.53) 63.43% male	0.25 (n=134)	Age 26.36 (9.01) 55.30% male	0.08 (n=302)
Numerical 2	Age 21.80 (3.19) 67.06% male	0.11 (n=252)	Age 24.84 (5.53) 60.83% male	-0.01 (n=577)
Abstract 2	Age 22.38 (5.11) 61.80% male	0.15 (n=102)	Age 26.30 (9.01) 57.67% male	0.08 (n=222)

Table 21: The association between UCAS points, degree class and PfS-Reasoning Tests

Despite methodological and measurement issues in the criterion-related validity studies, it is possible to draw some conclusions from the data. Importantly, it can be concluded that the PfS-Reasoning Tests are assessing constructs that are quite distinct from those assessed through established educational assessments. For test users, this means that the results from the PfS-Reasoning Tests provide information on respondents that is distinct from their educational attainments. As the nature of education changes, becoming more diverse in the types of courses offered and methods assessment, findings such as this support the benefits of psychometric testing in offering a level playing field for the fair assessment of abilities.

© 2003 – 2013 Team Focus Limited

The pattern of criterion-related data also indicates that the associations between the PfS-Reasoning Tests and academic attainment becomes less the further students progress through the education system. Coupled with the data from **Table 13**, showing that the correlations within the PfS-Reasoning Tests are lower in the higher level tests, this supports the view that abilities become more defined and distinct from each other as people mature and move through the education system. It also reflects education at younger years being more homogeneous (e.g. core GCSE subjects which all student study) compared to degrees-level courses which offer far a broader range of study and assessment options.

Usually, when new tests such as the PfS-Reasoning Tests are developed, only evidence for content and construct validity is available. Currently, further work on the criterion-related validity of the tests is being conducted, looking at how test scores relate to job performance. Ideally, all organisations that use assessments should conduct their own validity studies, examining the link between test scores and current or future job performance. This information allows the validity and value of testing within each organisation to be assessed.

The ability of tests to discriminate between test takers' abilities, and therefore their validity in selection and development decisions, can be increased by generating 'local norms'. The norms contained in this User's Guide refer to quite broad samples. Typically, applicants to any single organisation will be a far narrower group, as a result of both self-selection and selection by the organisation. Test scores from applicants and current employees can be used to generate norm tables specific to an organisation, thus allowing finer discrimination and a way of directly judging test performance in relation to other applicants. More information on the development of local norms can be obtained by contacting the test publisher.

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

References

Anderson, N. and Cunningham-Snell, N. (2000) Personnel selection. In Chmiel, N. (ed.) *Introduction to Work and Organizational Psychology: A European Perspective.* Oxford: Blackwell.

Baddeley, A.D., Hitch, G.J.L (1974). Working Memory, In G.A. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 8, pp. 47–89), New York: Academic Press.Blinkhorn, S. (1985) *Graduate and Managerial Assessment*. Windsor: nferNelson.

Bongard, M. M. (1972) *Pattern Recognition.* New York: Spartan Books.

CIPD (2000) *Recruitment: IPD Survey Report 14*. London: Institute of Personnel and Development.

CIPD (2006) *Recruitment, Retention and Turnover.* London: Institute of Personnel and Development.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

CollegeBoard (2003) 2003 College-Bound Seniors Tables and Related Items. Available onlint at: http://www.collegeboard.com/about/news_info/cbsenior/yr2003/html/links.html on 24.2.05

Freedle, R. O. (2003) Correcting the SAT's Ethnic and Social-Class Bias: A Method for Reestimating SAT Scores . *Harvard Educational Review 72* (3).

Great Britain. Statutes (1995) *Disability Discrimination Act. Chapter 50.* London: HMSO.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C. and Bentson, C. (1987) Meta-analysis of assessment centre validity. *Journal of Applied Psychology, 72*, 493-511.

Hunter, J. E. and Hunter, R. F. (1982) Validity and utility of alternative methods of job performance. *Psychological Bulletin, 96,* 72-98.

Jenkins, A. (2001) *Companies' Use of Psychometric Testing and the Changing Demand for Skills: A Review of the Literature.* London: London School of Economics and Political Science.

Kyllonen, P. C. and Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389-433.

© 2003 – 2013 Team Focus Limited

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L. and Maurer, S. D. (1994) The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79,* 599-616.

Messick, S. (1995) Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific enquiry into score meaning. *American Psychologist, 50,* 741-749.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101.

Reilly, R. R. and Chao, G. T. (1982) Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35,* 1-62.

Robertson, I. T. and Kinder, A. (1993) Personality and job competencies: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology, 66,* 225-244.

Shackleton, V. and Newell, S (1991) Management selection: A comparative survey of methods used in top British and French companies. *Journal of Occupational Psychology, 64,* 23-36.

Schmidt, F. L. and Hunter, J. E. (1998) The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-274.

© 2003 – 2013 Team Focus Limited

Appendix One: Explanations of practice questions

Note: The explanations of the practice questions given here relate only to the closed versions of the PfS-Reasoning Tests. Full explanations of the practice questions on the open versions are displayed to test takers by the computer, after they have given an answer to each practice question.

Verbal Reasoning

Levels 1 and 2

P1: Modern methods of predicting the weather are not always accurate.

The correct answer to this statement is 'can't tell'. Although we know that weather forecasts are not always accurate, the passage gives no information about how accurate modern methods of predicting the weather are. As no information on the accuracy of modern methods is given in the passage, we do not know whether this statement is true or not.

P2: Personal observations can be accurate predictors of the weather.

This statement is 'true' as the passage states that 'Before modern weather forecasts, people relied on their own observations to predict the weather'. It also says that the 'red sky' rhyme that came from these observations 'is quite a good indicator of what the weather is going to be.' Therefore, the weather can be accurately predicted from personal observations.

P3: If there is a 'red sky' in the morning, there is a good chance that the weather will be fine.

The answer to this statement is 'false'. The rhyme 'red sky in the morning, shepherd's warning' tells us to expect 'bad weather' and we are told that ''red sky' is quite a good indicator of what the weather is going to be'. Therefore, a red sky in the morning is likely to indicate bad weather.

P4: All weather rhymes are poor predictors of the weather.

This statement is 'false'. The passage says that the 'red sky' rhyme is 'quite a good indicator of what the weather is going to be', so not all rhymes are poor predictors of the weather.

© 2003 – 2013 Team Focus Limited

Levels 3 and 4

P1: Modern methods of predicting the weather are not always accurate.

The correct answer to this statement is ‘can’t tell’. Although it is commonly known that weather forecasts are not always accurate, the passage gives information about the accuracy of proverbs in predicting weather but not the accuracy of modern prediction methods. As no information on the accuracy of modern methods is given in the passage, the truth of this statement is not known.

P2: Atmospheric conditions can indicate what the weather is likely to be.

This statement is ‘true’ as the passage states that ‘It was known that certain atmospheric conditions were likely to lead to different types of weather’. Although many of the proverbs which came from observations of atmospheric conditions ‘predict the weather no better than chance’, the ‘red sky’ proverb is ‘quite an accurate predictor of the weather’, indicating that the weather can be predicted from atmospheric conditions.

P3: All proverbs are poor predictors of the weather.

This statement is ‘false’. The passage says that ‘many proverbs have been shown to predict the weather no better than chance, but the ‘red sky’ proverb is ‘quite an accurate predictor of the weather’. Therefore not all proverbs are poor predictors as the statement suggests.

P4: If there is a ‘red sky’ in the morning, there is a good chance that the weather will be fine.

The answer to this statement is ‘false’. The proverb ‘red sky in the morning, shepherd’s warning’ tells us to expect ‘bad weather’ and we are told that ‘‘red sky’ is quite an accurate predictor of the weather’. Therefore, a red sky in the morning is likely to indicate bad weather.

© 2003 – 2013 Team Focus Limited

Numerical Reasoning

Levels 1 and 2

P1: How many employees does the company have altogether?

The correct answer is 50. This is found by adding the numbers in the 'Number of employees' column.

P2: How many employees travel 8 miles or more to work?

The correct answer is 'Can't tell'. Although you are told that 15 employees travel between 6 and 10 miles to work, you cannot tell how many of these 15 employees travel 8 or more miles.

P3: Which is the most common distance that employees travel?

The correct answer is 1 to 5 miles. This is the distance travelled to work by most employees (17).

P4: What percentage of employees travel between 1 and 5 miles to work?

The correct answer is 34%. To find this you need to divide the number of employees who travel between 1 and 5 miles to work (17) by the total number of employees (50) and multiply this figure by 100 to give you a percentage.

© 2003 – 2013 Team Focus Limited

Levels 3 and 4

P1: In which year did rural houses show their greatest change in value?

The correct answer is Year 3, as the graph shows that rural house prices increased by 8% in Year 3.

P2: In which year was the greatest difference between the change in the value of houses in rural and urban areas?

The correct answer is Year 5. To find the difference in the change of rural and urban house prices you have to subtract the smaller value for each year from the larger value. The largest difference (7%) is for Year 5.

P3: A house in an urban area is worth £110,000 at the beginning of Year 1. What is it likely to be worth at the end of Year 2?

The correct answer is £106,722. The graph shows that houses in urban areas lost 1% of their value in Year 1 (£1100 on a house worth £110,000) so the value of the house at the end of Year 1 is £108,900. In year 2, 2% of the value is lost (£2178 on a house worth £108,900), making the value £106,722.

P4: In which year did the combined value of rural and urban houses change the most?

The correct answer is 'Can't tell'. It is not possible to tell in which year the combined value of rural and urban houses changed the most, without knowing the proportion of houses that are classified as being 'rural' and 'urban' and the average value each year.

© 2003 – 2013 Team Focus Limited

Abstract Reasoning

Levels 1 and 2

P1: The correct answer is ‘Set A’. All the shapes in Set A have three triangles. Two of the triangles point upwards and are white, the other points downwards and is black. All the shapes in Set B have three diamonds.

P2: The correct answer is ‘Neither’, as all the shapes in Set A have two white triangles pointing upwards and one black triangle pointing downwards, and all the shapes in Set B have three diamonds.

P3: The correct answer is ‘Set B’. All the shapes in Set A have three triangles. Two of the triangles point upwards and are white, the other points downwards and is black. All the shapes in Set B have three diamonds.

P4: The correct answer is ‘Neither’, as all the shapes in Set A have two white triangles pointing upwards and one black triangle pointing downwards, and all the shapes in Set B have three diamonds.

P5: The correct answer is ‘Set B’. All the shapes in Set A have three triangles. Two of the triangles point upwards and are white, the other points downwards and is black. All the shapes in Set B have three diamonds.

© 2003 – 2013 Team Focus Limited

Levels 3 and 4

P1: The correct answer is ‘Set A’. All the shapes in Set A have at least one white triangle. As this is the only common feature in Set A, all other features should be ignored. All the shapes in Set B have at least one black square. Again, as this is the only common feature in Set B, all other features should be ignored.

P2: The correct answer is ‘Neither’, as all the shapes in Set A have at least one white triangle and all the shapes in Set B have at least one black square.

P3: The correct answer is ‘Set B’. All the shapes in Set A have at least one white triangle. As this is the only common feature in Set A, all other features should be ignored. All the shapes in Set B have at least one black square. Again, as this is the only common feature in Set B, all other features should be ignored.

P4: The correct answer is ‘Neither’, as all the shapes in Set A have at least one white triangle and all the shapes in Set B have at least one black square.

P5: The correct answer is ‘Set B’. All the shapes in Set A have at least one white triangle. As this is the only common feature in Set A, all other features should be ignored. All the shapes in Set B have at least one black square. Again, as this is the only common feature in Set B, all other features should be ignored.

© 2003 – 2013 Team Focus Limited

Appendix Two: Sample test reports

www.profilingforsuccess.com

Administrator's Report

Numerical Reasoning - Level 3

Denise Debutante

Test Results

This report describes Denise Debutante's results on Level 3 of the Numerical Reasoning Test. This test assesses the ability to use numerical information to solve problems.

On the Numerical Reasoning Test, Denise attempted 23 questions out of 36 and answered 15 correctly. To put this raw score into context, it has been compared with the following group: Undergraduate students (n=761).

In relation to the comparison group, Denise's scores are as follows:

T-Score	45
68% T-Score confidence band	41-48
80% T-Score confidence band	40-49
Percentile	32

The T-Score and 68% T-Score confidence band are shown below:

T-Score	20	30	40	50	60	70	80
Percentile	1	2	16	50	84	98	99

Speed and Accuracy

Combining information on the number of questions attempted and the number answered correctly, indicates that Denise attempted as many questions as the majority of the comparison group. Of the questions attempted, she answered correctly an average number.

Overall, this pattern of responses was similar to most others in the comparison group. Denise worked at a similar rate to others and tended to answer correctly a similar proportion of the questions she attempted. She could be seen to have achieved a balance between speed and accuracy but has done so to an average extent, leaving room for improvement in performance. There is no indication that she sacrificed accuracy for speed or vice versa.

To perform better Denise would have to work more quickly and more accurately. It would be worth exploring whether this balance of speed and accuracy is typical of how Denise approaches her work and other activities, or whether it reflects a style adopted just for the test. In some situations it can be valuable to vary the approach taken; sometimes placing greater emphasis on accuracy and at other times emphasising speed. As Denise answered correctly only an average number of questions she attempted, to improve her performance she could improve her accuracy, and then try to work more quickly.

© 2003 – 2013 Team Focus Limited

Suggested Review Prompts

- How do you feel about the Numerical Reasoning Test?
- Have you taken this type of test before, and if so, how did you find it?
- What parts of the test did you find most challenging?
- How did you feel when you were doing the test?
- Your approach to the test seemed to be as fast and accurate as most of the comparison group. To what extent is this characteristic of your working style generally?
- Your test results suggest an average level of speed and accuracy in relation to the comparison group. Think of times when you have been able to, or your work has required you to, work quickly and possibly sacrifice some of your accuracy, or vice versa? How did you feel when you had to work like this?
- As you answered correctly only an average number of the questions you attempted, to perform better you would need to improve your accuracy, and then possibly your speed. What activities would you enjoy or be willing to do in order to practise the kinds of skills needed for the Numerical Reasoning Test?
- If you were to take the test again, how would you approach it differently?

Notes On Interpreting This Report

When reading this report, you should remember that:

- psychometric tests are only one source of information about a person's abilities and style, so results should be integrated with other evidence to provide as broad a picture as possible. How much the test results will influence any final assessment will depend on the appropriateness of the tests and the quality of the other information collected.
- all test scores (as with any measurement) are subject to error. Scores are therefore taken as an indication of the band of ability within which the individual might fall.
- scores may change due to error and small differences between scores may not be significant. The amount of error can be estimated statistically and this is how the range of scores quoted in this report has been determined.
- high scores are easier to interpret than low scores. If people score highly, then they probably do have a high level of the ability in question. People can, however, get low scores for many reasons - misunderstanding, lack of familiarity with test procedures, anxiety, etc. Low scores should therefore be interpreted as 'the individual has not yet shown evidence of this ability'.
- all scores are compared to groups of individuals, e.g. people at various stages of their education, those working in different jobs. Therefore the score is not fixed. A score may be above average compared to one group and below average compared to another.
- the results show how the person performed on the test on this particular occasion. A person's score is likely to fluctuate according to a number of different factors: this means that scores might differ slightly if the test were taken on a second occasion.
- the test results are an opportunity for the individual to demonstrate their abilities with particular types of reasoning and problem solving. They do not cover all kinds of reasoning. However, psychometric tests, properly chosen, have been found to contribute usefully to an overall assessment of an individual's abilities. They must be properly integrated with other data and should never be used on their own.

ID/REF: ABX-470; email: denise@debutantesanon.co.uk
Job title = Account Executive; Department = Software Sales
Date tested: 23/2/2007 Norm used: Undergraduate students (n=761)
Copyright © Profiling for Success 2007
Profiling for Success is published by Team Focus Limited.

© 2003 – 2013 Team Focus Limited

Administrator's Summary Report

Numerical Reasoning - Level 3

Denise Debutante

Test Results

This report describes Denise Debutante's results on Level 3 of the Numerical Reasoning Test. This test assesses the ability to use numerical information to solve problems.

On the Numerical Reasoning Test, Denise attempted 23 questions out of 36 and answered 15 correctly. To put this raw score into context, it has been compared with the following group: Undergraduate students (n=761).

In relation to the comparison group, Denise's scores are as follows:

T-Score	45
68% T-Score confidence band	41-48
80% T-Score confidence band	40-49
Percentile	32

The T-Score and 68% T-Score confidence band are shown below:

T-Score	20	30	40	50	60	70	80
Percentile	1	2	16	50	84	98	99

Speed and Accuracy

Combining information on the number of questions attempted and the number answered correctly indicates that Denise attempted as many questions as the majority of the comparison group. Of the questions attempted, she answered correctly an average number.

Notes On Interpreting This Report

When reading this report, you should remember that test results are only one source of information about a person's abilities, and the test Denise has taken looks at a very specific type of ability. All test scores are subject to error and so scores indicate a band of ability within which the test taker might fall, so an obtained score may under or over estimate ability. Low test scores can occur for many reasons - misunderstanding, lack of familiarity, anxiety - and the score may change if the test is taken again.

ID/REF: ABX-470; email: denise@debutantesanon.co.uk
Job title = Account Executive; Department = Software Sales
Date tested: 23/2/2007 Norm used: Undergraduate students (n=761)
Copyright © Profiling for Success 2007
Profiling for Success is published by Team Focus Limited.

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Feedback Report

Numerical Reasoning - Level 3

Denise Debutante

Your Results

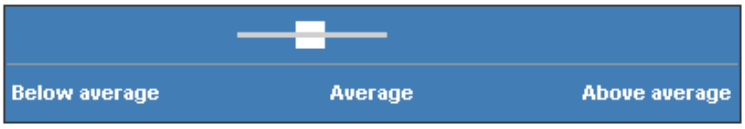
This report describes your results on the Numerical Reasoning Test which looks at your ability to use numerical information to solve problems.

On the Numerical Reasoning Test you attempted 23 of the 36 questions in the test, and answered 15 of these correctly.

To put your score into context, it is compared to a large group of people who have already taken the test. In this case, your test score has been compared to the following group: Undergraduate students (n=761)

Your results are shown graphically below. The small square indicates the score you obtained on the Numerical Reasoning Test in relation to the comparison group. However, as measurement is never totally accurate the line passing through the small square shows the range which gives the best estimate of you ability on the test

When compared to the comparison group, your score was at the 32nd percentile. This means you scored as well as 32 per cent of the comparison group.



Possible Reasons For Your Performance

This section of the report combines information on how quickly you worked at the test and how accurate your answers were.

In the time allowed for the test you attempted an average number of questions and answered an average number of these correctly. This pattern of performance suggests that you:

- appear to have understood what the test required you to do.
- seem to have achieved a reasonable balance between speed and accuracy, but there is still room for improvement.

You may like to consider the following points, some of which may help you to improve your performance if you were to take a test like this again:

- To improve your performance on tests like this you would need to improve both your accuracy and speed of working.

© 2003 – 2013 Team Focus Limited

- Ways to improve your accuracy could include; reading each question more carefully and making sure you understand what you are being asked to do; thinking about how to go about answering the question; making sure that you have read the details in the question accurately.

- Ways to improve your speed could include; making sure you focus on the test and that you are not distracted; skipping any questions you get stuck on; spending less time double-checking answers you are pretty sure of and more time on questions you find difficult.

- Your approach to the test seemed to be as fast and as accurate as most of the comparison group. To what extent is this characteristic of your working style generally?

- Think of some activities you would enjoy or be willing to do in order to practise the kinds of skills needed for the test.

- If you were to take the test again, how would you approach it differently?

Notes On Interpreting This Report

When reading this report, the following points should be considered:

- psychometric tests are only one source of information about your abilities and style, and the test you have taken looks at a very specific type of ability. However, tests are known to be a useful part of an overall assessment of a person's abilities.

- all test scores (as with any measurement) are subject to error. The scores therefore indicate a band of ability within which you might fall, so your obtained score may under or over estimate your ability.

- high scores are easier to interpret than low scores. If people score highly, they are likely to have the ability being measured. People can, however, get low scores for many reasons – misunderstanding, lack of familiarity with tests, anxiety, etc. Low scores should therefore be seen as showing 'you have not yet shown evidence of this ability on this test'.

- all scores are compared to groups of individuals, e.g. people at various stages of their education, those working in different jobs. Therefore the score is not fixed. A score may be above average compared to one group and below average compared to another.

- the results show how you performed on the test on this particular occasion. Your score can fluctuate according to a number of different factors: this means that your score may change if you took the test again.

Date tested: 23/2/2007 Norm used: Undergraduate students (n=761)
Copyright © Profiling for Success 2007
Profiling for Success is published by Team Focus Limited.

Feedback Summary Report

Numerical Reasoning - Level 3

Denise Debutante

Your Results

This report describes your results on the Numerical Reasoning Test which looks at your ability to use numerical information to solve problems.

On the Numerical Reasoning Test you attempted 23 of the 36 questions in the test, and answered 15 of these correctly.

To put your score into context, it is compared to a large group of people who have already taken the test. In this case, your test score has been compared to the following group: Undergraduate students (n=761)

Your results are shown graphically below. The small square indicates the score you obtained on the Numerical Reasoning Test in relation to the comparison group. However, as measurement is never totally accurate the line passing through the small square shows the range which gives the best estimate of you ability on the test

When compared to the comparison group, your score was at the 32nd percentile. This means you scored as well as 32 per cent of the comparison group.

The Numerical Reasoning Test also looks at how quickly you worked at the test and how accurate your answers were.

In the time allowed for the test you attempted an average number of questions and answered correctly an average number of these.

Notes On Interpreting This Report

When reading this report, you should remember that test results are only one source of information about your abilities, and the test you have taken looks at a very specific type of ability. All test scores are subject to error and scores indicate a band of ability within which you might fall, so your obtained score may under or over estimate your ability. Low test scores can occur for many reasons - misunderstanding, lack of familiarity, anxiety - and your score may change if you take the test again.

Date tested: 23/2/2007 Norm used: Undergraduate students (n=761)
Copyright © Profiling for Success 2007
Profiling for Success is published by Team Focus Limited.

© 2003 – 2013 Team Focus Limited

Appendix Three: Norm tables

Introduction to the norm tables

This appendix contains the norm tables for the PfS-Reasoning Tests. For the tests that are available in paper-and-pencil format – Verbal, Numerical and Abstract closed tests levels 1 to 4 – full norm tables and given. For the tests that are only available through the PfS online assessments system – Verbal, Numerical and Abstract open tests levels 1 and 2, and the Combined Reasoning Test – descriptions of the norms available are given, but not the full norm tables. Full norm tables are not given for the computer-based tests as all necessary comparisons are done by the PfS assessment system on submission of test results, and given in the test reports (see Appendix Two for sample reports).

The first part of this appendix contains the norm tables that were constructed from the initial standardisations of the tests. The second part gives updates to the initial norm tables and norms for more specific groups that have collected from specific organisations use of the PfS-Reasoning Tests. Further updates will be added as new norms become available.

Norm tables available for the PfS-Reasoning Tests

© 2003 – 2013 Team Focus Limited

General norms for closed tests

Test: Verbal Reasoning Level: 1

Description of norm group: GCSE students and students in their first year of courses at FE institutions. Young people on vocational training courses and employees in basic level jobs.

Size of norm group	210	Reliability	0.90
Mean	16.62	SEM (raw scores)	1.82
SD	5.73	SED (68/80/95%; raw scores)	2.57 / 3.29 / 5.15

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
30 - 32	99	74	73 - 76	72 - 77
29	98	71	70 - 73	69 - 74
28	98	70	68 - 71	67 - 72
27	97	69	67 - 70	66 - 71
26	96	68	66 - 69	65 - 70
25	94	66	64 - 68	63 - 68
24	91	63	62 - 65	61 - 66
23	87	61	59 - 63	59 - 64
22	82	59	57 - 61	57 - 62
21	76	57	55 - 59	55 - 59
20	69	55	53 - 57	52 - 57
19	61	53	51 - 55	51 - 55
18	55	51	49 - 53	49 - 54
17	49	50	48 - 52	48 - 52
16	44	49	47 - 50	46 - 51
15	40	47	46 - 49	45 - 50
14	35	46	44 - 48	44 - 49
13	31	45	43 - 47	43 - 47
12	26	44	42 - 45	41 - 46
11	22	42	40 - 44	40 - 44
10	16	40	38 - 42	38 - 42
9	10	37	35 - 39	35 - 40
8	6	34	32 - 36	32 - 37
7	3	32	30 - 34	29 - 34
6	2	30	28 - 31	27 - 32
0 - 5	1	26	25 28	24 29

© 2003 – 2013 Team Focus Limited

Test: Verbal Reasoning Level: 2

Description of norm group: FE students studying a range of vocational and academic courses, at institutions predominantly in the south east of England. A limited number of currently employed people were also included in this sample.

Size of norm group	303	Reliability	0.80
Mean	16.32	SEM (raw scores)	2.32
SD	5.18	SED (68/80/95%; raw scores)	3.28 / 4.19 / 6.56

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
27 - 32	99	74	71 - 76	71 - 77
26	98	71	68 - 73	68 - 74
25	96	67	65 - 70	64 - 70
24	92	64	62 - 67	61 - 67
23	89	62	60 - 64	59 - 65
22	84	60	57 - 62	57 - 63
21	77	58	55 - 60	55 - 61
20	71	56	53 - 58	53 - 59
19	67	54	52 - 57	51 - 57
18	63	53	51 - 56	50 - 56
17	57	52	49 - 54	49 - 55
16	50	50	48 - 52	47 - 53
15	42	48	46 - 50	45 - 51
14	35	46	44 - 49	43 - 49
13	29	45	42 - 47	42 - 48
12	24	43	40 - 45	40 - 46
11	17	41	38 - 43	38 - 44
10	11	38	36 - 40	35 - 41
9	7	36	33 - 38	33 - 39
8	5	33	31 - 36	30 - 36
7	3	31	29 - 33	28 - 34
0 - 6	1	28	26 - 30	25 - 31

© 2003 – 2013 Team Focus Limited

Test: Verbal Reasoning Level: 3

Description of norm group: Undergraduate students from a range of universities including 'old' institutions (e.g. London), 'redbrick' institutions (e.g. Derby, Sussex) and 'new' universities (e.g. Uxbridge, Brighton). This sample also included a number of people currently employed in a range of positions.

Size of norm group	1322	Reliability	0.86
Mean	24.10	SEM (raw scores)	2.27
SD	6.07	SED (68/80/95%; raw scores)	3.21 / 4.11 / 6.42

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
35 - 40	99	73	70 - 75	70 - 76
34	97	69	67 - 71	66 - 72
33	95	66	64 - 69	64 - 69
32	92	64	62 - 66	61 - 67
31	88	62	59 - 64	59 - 65
30	83	59	57 - 62	57 - 62
29	77	57	55 - 60	54 - 60
28	71	55	53 - 58	53 - 58
27	65	54	51 - 56	51 - 57
26	58	52	50 - 54	49 - 55
25	52	50	48 - 53	48 - 53
24	46	49	47 - 51	46 - 52
23	40	47	45 - 50	44 - 50
22	30	46	44 - 48	43 - 49
21	30	45	42 - 47	42 - 48
20	25	43	41 - 46	40 - 46
19	21	42	40 - 44	39 - 45
18	17	40	38 - 43	37 - 43
17	11	39	37 - 41	36 - 42
16	11	37	35 - 40	35 - 40
15	8	36	34 - 38	33 - 39
14	7	35	33 - 37	32 - 38
13	5	33	31 - 36	31 - 36
12	4	32	30 - 34	29 - 35
11	3	31	29 - 33	28 - 34
10	2	30	28 - 32	27 - 33
9	2	29	27 - 31	26 - 32
0 - 8	1	28	25 - 30	25 - 31

© 2003 – 2013 Team Focus Limited

Test: Verbal Reasoning Level: 4

Description of norm group: Postgraduate students and experienced professionals. Some undergraduates from established universities (e.g. London, Reading, Sussex).

Size of norm group	1131	Reliability	0.87
Mean	25.45	SEM (raw scores)	2.26
SD	6.27	SED (68/80/95%; raw scores)	3.20 / 4.09 / 6.39

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
37 – 40	99	76	73 - 78	73 - 79
36	98	71	69 - 74	68 - 74
35	96	68	66 - 70	65 - 71
34	93	65	63 - 67	62 - 68
33	90	63	60 - 65	60 - 65
32	85	60	58 - 62	57 - 63
31	79	58	56 - 60	55 - 61
30	73	56	54 - 58	53 - 59
29	67	54	52 - 57	51 - 57
28	60	53	50 - 55	50 - 56
27	54	51	49 - 53	48 - 54
26	49	50	47 - 52	47 - 53
25	44	48	46 - 51	46 - 51
24	40	47	45 - 50	44 - 50
23	34	46	44 - 48	43 - 49
22	29	44	42 - 47	42 - 47
21	24	43	41 - 45	40 - 46
20	20	42	39 - 44	39 - 44
19	17	40	38 - 43	37 - 43
18	14	39	37 - 42	36 - 42
17	12	38	36 - 40	35 - 41
16	9	37	34 - 39	34 - 40
15	7	35	33 - 38	32 - 38
14	5	34	32 - 36	31 - 37
13	4	32	30 - 35	30 - 35
12	3	31	28 - 33	28 - 33
11	2	29	26 - 31	26 - 32
0 – 10	1	27	25 - 30	24 - 30

© 2003 – 2013 Team Focus Limited

Test: Numerical Reasoning **Level: 1**

Description of norm group: GCSE students and students in their first year of courses at FE institutions. Young people on vocational training courses and employees in basic level jobs.

Size of norm group	250	Reliability	0.93
Mean	19.30	SEM (raw scores)	1.23
SD	4.64	SED (68/80/95%; raw scores)	1.73 / 2.21 / 3.45

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
28	99	75	74 - 76	73 - 76
27	97	69	68 - 70	68 - 71
26	94	66	64 - 67	64 - 67
25	90	63	62 - 64	61 - 64
24	84	60	59 - 61	59 - 62
23	78	58	56 - 59	56 - 59
22	70	55	54 - 56	54 - 57
21	62	53	52 - 54	51 - 55
20	53	51	49 - 52	49 - 52
19	43	48	47 - 49	47 - 50
18	34	46	45 - 47	44 - 48
17	27	44	43 - 45	42 - 45
16	21	42	41 - 43	40 - 43
15	16	40	39 - 41	38 - 42
14	13	39	37 - 40	37 - 40
13	10	37	36 - 39	36 - 39
12	8	36	35 - 37	34 - 37
11	6	35	33 - 36	33 - 36
10	5	33	32 - 34	31 - 35
9	3	31	30 - 32	30 - 33
8	2	30	28 - 31	28 - 31
7	2	29	28 - 30	27 - 30
0 – 6	1	27	26 - 29	26 - 29

© 2003 – 2013 Team Focus Limited

Test: Numerical Reasoning **Level: 2**

Description of norm group: FE students studying a range of vocational and academic courses, at institutions predominantly in the south east of England. A limited number of currently employed people were also included in this sample.

Size of norm group	337	Reliability	0.84
Mean	14.95	SEM (raw scores)	1.90
SD	4.74	SED (68/80/95%; raw scores)	2.69 / 3.44 / 5.37

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
26 – 28	99	73	71 - 75	70 - 75
25	97	69	68 - 71	67 - 72
24	96	67	65 - 69	65 - 69
23	94	65	63 - 67	63 - 68
22	91	64	62 - 65	61 - 66
21	88	62	60 - 64	59 - 64
20	83	60	58 - 62	57 - 62
19	78	58	56 - 60	55 - 60
18	74	56	54 - 58	54 - 59
17	69	55	53 - 57	52 - 57
16	63	53	51 - 55	51 - 56
15	55	51	49 - 53	49 - 54
14	46	49	47 - 51	47 - 51
13	37	47	45 - 49	44 - 49
12	28	44	42 - 46	42 - 47
11	20	42	40 - 44	39 - 44
10	15	39	38 - 41	37 - 42
9	10	37	35 - 39	35 - 40
8	6	35	33 - 37	32 - 37
7	3	32	30 - 34	29 - 34
0 – 6	1	28	26 - 30	26 - 30

© 2003 – 2013 Team Focus Limited

Test: Numerical Reasoning Level: 3

Description of norm group: Undergraduate students from a range of universities including 'old' institutions (e.g. London), 'redbrick' institutions (e.g. Derby, Sussex) and 'new' universities (e.g. Uxbridge, Brighton). This sample also included a number of people currently employed in a range of positions.

Size of norm group	1609	Reliability	0.87
Mean	18.04	SEM (raw scores)	2.05
SD	5.69	SED (68/80/95%; raw scores)	2.90 / 3.71 / 5.80

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
32 – 36	99	72	70 - 74	70 - 75
31	98	70	68 - 72	68 - 73
30	97	69	67 - 71	66 - 71
29	96	67	65 - 69	64 - 70
28	94	66	63 - 68	63 - 68
27	92	64	62 - 66	62 - 67
26	91	63	61 - 65	60 - 66
25	88	62	60 - 64	59 - 65
24	86	61	59 - 63	58 - 63
23	82	59	57 - 61	57 - 62
22	77	57	55 - 59	55 - 60
21	71	56	54 - 58	53 - 58
20	65	54	52 - 56	51 - 57
19	59	52	50 - 54	50 - 55
18	53	51	49 - 53	48 - 53
17	46	49	47 - 51	46 - 52
16	39	47	45 - 49	45 - 50
15	31	45	43 - 47	42 - 48
14	24	43	41 - 45	40 - 45
13	18	41	39 - 43	38 - 43
12	13	39	37 - 41	36 - 41
11	10	37	35 - 39	34 - 40
10	7	35	33 - 37	33 - 38
9	5	33	31 - 36	31 - 36
8	3	32	30 - 34	29 - 34
7	2	30	28 - 32	27 - 33
0 – 6	1	27	25 - 29	25 - 30

© 2003 – 2013 Team Focus Limited

Test: Numerical Reasoning **Level: 4**

Description of norm group: Postgraduate students and experienced professionals. Some undergraduates from established universities (e.g. London, Reading, Sussex).

Size of norm group	1510	Reliability	0.89
Mean	16.24	SEM (raw scores)	2.16
SD	6.50	SED (68/80/95%; raw scores)	3.05 / 3.91 / 6.11

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
33 – 36	99	72	70 - 75	70 - 75
32	98	71	69 - 73	68 - 74
31	98	70	68 - 72	67 - 72
30	97	68	66 - 71	66 - 71
29	96	67	65 - 69	64 - 70
28	94	66	63 - 68	63 - 68
27	92	64	62 - 66	62 - 67
26	91	63	61 - 65	60 - 66
25	89	62	60 - 64	59 - 65
24	86	61	59 - 63	58 - 64
23	84	60	58 - 62	57 - 63
22	81	59	57 - 61	56 - 61
21	78	58	55 - 60	55 - 60
20	74	57	54 - 59	54 - 59
19	71	55	53 - 58	53 - 58
18	66	54	52 - 56	51 - 57
17	60	52	50 - 55	50 - 55
16	53	51	49 - 53	48 - 54
15	47	49	47 - 51	47 - 52
14	41	48	46 - 50	45 - 51
13	35	46	44 - 48	43 - 49
12	29	44	42 - 47	42 - 47
11	23	42	40 - 45	40 - 45
10	17	41	38 - 43	38 - 43
9	13	39	36 - 41	36 - 41
8	9	36	34 - 39	34 - 39
7	5	34	32 - 36	31 - 37
6	3	31	29 - 33	28 - 34
0 – 5	1	26	24 - 29	24 - 29

© 2003 – 2013 Team Focus Limited

Test: Abstract Reasoning **Level: 1**

Description of norm group: GCSE students and students in their first year of courses at FE institutions. Young people on vocational training courses and employees in basic level jobs.

Size of norm group	156	Reliability	0.93
Mean	28.51	SEM (raw scores)	2.07
SD	7.82	SED (68/80/95%; raw scores)	2.93 / 3.75 / 5.86

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
44 – 50	99	72	70 - 75	70 - 75
43	98	70	68 - 72	67 - 73
42	97	69	67 - 71	66 - 71
41	96	67	65 - 69	65 - 70
40	94	65	63 - 67	63 - 68
39	91	63	61 - 66	61 - 66
38	89	62	60 - 64	59 - 65
37	86	61	59 - 63	58 - 64
36	83	60	58 - 62	57 - 62
35	79	58	56 - 60	55 - 61
34	74	56	54 - 58	54 - 59
33	68	55	53 - 57	52 - 57
32	64	54	52 - 56	51 - 56
31	61	53	51 - 55	50 - 55
30	57	52	50 - 54	49 - 54
29	51	50	48 - 52	48 - 53
28	44	49	47 - 51	46 - 51
27	40	47	45 - 49	45 - 50
26	36	46	44 - 48	44 - 49
25	32	45	43 - 47	43 - 48
24	29	44	42 - 46	42 - 47
23	25	43	41 - 45	41 - 46
22	21	42	40 - 44	39 - 45
21	17	40	38 - 42	38 - 43
20	14	39	37 - 41	36 - 42
19	12	38	36 - 40	35 - 41
18	10	37	35 - 39	35 - 40
17	8	36	34 - 38	33 - 39
16	7	35	33 - 37	33 - 38
15	6	34	32 - 36	32 - 37
14	5	33	31 - 35	31 - 36
13	3	32	30 - 34	29 - 34
12	3	31	28 - 33	28 - 33
11	2	30	28 - 32	27 - 32
0 – 10	1	28	26 - 30	25 - 30

© 2003 – 2013 Team Focus Limited

Test: Abstract Reasoning **Level: 2**

Description of norm group: FE students studying a range of vocational and academic courses, at institutions predominantly in the south east of England. A limited number of currently employed people were also included in this sample.

Size of norm group	242	Reliability	0.87
Mean	20.80	SEM (raw scores)	2.97
SD	8.24	SED (68/80/95%; raw scores)	4.20 / 5.38 / 8.40

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
43 – 50	99	72	69 - 75	68 - 76
42	98	71	68 - 73	67 - 74
41	98	70	67 - 73	66 - 74
40	98	70	67 - 73	66 - 74
39	97	69	66 - 72	65 - 73
38	97	68	65 - 71	64 - 72
37	96	67	64 - 70	63 - 71
36	95	67	64 - 69	63 - 70
35	95	66	63 - 69	62 - 70
34	94	66	63 - 69	62 - 70
33	92	64	61 - 67	61 - 68
32	90	63	60 - 66	59 - 66
31	87	61	58 - 64	57 - 65
30	85	60	57 - 63	56 - 64
29	83	60	57 - 63	56 - 63
28	81	59	56 - 62	55 - 63
27	79	58	55 - 61	54 - 62
26	76	57	54 - 60	53 - 61
25	72	56	53 - 59	52 - 60
24	69	55	52 - 58	51 - 59
23	64	54	51 - 57	50 - 57
22	59	52	49 - 55	48 - 56
21	55	51	48 - 54	47 - 55
20	51	50	47 - 53	46 - 54
19	46	49	46 - 52	45 - 53
18	41	48	45 - 51	44 - 52
17	37	47	44 - 50	43 - 50
16	32	45	42 - 48	42 - 49
15	27	44	41 - 47	40 - 48
14	23	42	39 - 45	39 - 46
13	19	41	38 - 44	37 - 45
12	15	40	37 - 43	36 - 43
11	11	38	35 - 41	34 - 42
10	8	36	33 - 39	32 - 40
9	5	34	31 - 37	30 - 37
8	2	30	27 - 33	26 - 33
0 – 7	1	21	18 - 24	17 - 25

© 2003 – 2013 Team Focus Limited

Test: Abstract Reasoning **Level: 3**

Description of norm group: Undergraduate students from a range of universities including 'old' institutions (e.g. London), 'redbrick' institutions (e.g. Derby, Sussex) and 'new' universities (e.g. Uxbridge, Brighton). This sample also included a number of people currently employed in a range of positions.

Size of norm group	860	Reliability	0.92
Mean	31.2	SEM (raw scores)	3.16
SD	11.18	SED (68/80/95%; raw scores)	4.47 / 5.72 / 8.94

Raw score	Percentile rank	T-score	68% T-score confidence band		80% T-score confidence band	
54 – 60	99	72	69	75	68 -	76
53	98	71	67	74	67 -	75
52	97	69	66	72	65 -	73
51	96	68	65	71	64 -	72
50	96	67	64	70	63 -	71
49	95	66	63	69	62 -	70
48	93	65	62	68	61 -	69
47	91	64	60	67	60 -	68
46	90	63	59	66	58 -	67
45	88	62	58	65	58 -	66
44	86	61	57	64	57 -	65
43	84	60	57	63	56 -	64
42	81	59	56	62	55 -	63
41	80	58	55	61	54 -	62
40	77	57	54	61	53 -	61
39	75	57	53	60	53 -	61
38	72	56	53	59	52 -	60
37	70	55	52	58	51 -	59
36	66	54	51	57	50 -	58
35	62	53	50	56	49 -	57
34	59	52	49	55	48 -	56
33	55	51	48	54	47 -	55
32	52	50	47	54	46 -	54
31	48	50	46	53	46 -	54
30	46	49	46	52	45 -	53
29	43	48	45	51	44 -	52
28	40	48	44	51	44 -	52
27	37	47	44	50	43 -	51
26	34	46	43	49	42 -	50

table continued overleaf…

© 2003 – 2013 Team Focus Limited

Abstract 3 norm table continued

25	31	45	42	48	41	-	49
24	28	44	41	47	40	-	48
23	25	43	40	46	39	-	47
22	22	42	39	46	38	-	46
21	20	42	38	45	37	-	46
20	17	41	37	44	37	-	45
19	15	40	36	43	36	-	44
18	12	38	35	42	34	-	43
17	10	37	34	40	33	-	41
16	8	36	33	39	32	-	40
15	7	35	32	39	31	-	39
14	6	35	32	38	31	-	39
13	6	34	31	37	30	-	38
12	4	33	30	36	29	-	37
11	3	31	28	35	27	-	35
10	2	30	27	33	26	-	34
9	2	30	27	33	26	-	34
8	2	29	26	32	25	-	33
7	2	29	26	32	25	-	33
0 - 6	1	28	24	31	24	-	32

© 2003 – 2013 Team Focus Limited

Test: Abstract Reasoning **Level: 4**

Description of norm group: Postgraduate students and experienced professionals. Some undergraduates from established universities (e.g. London, Reading, Sussex).

Size of norm group	881	Reliability	0.91
Mean	30.35	SEM (raw scores)	3.13
SD	10.41	SED (68/80/95%; raw scores)	4.43 / 5.67 / 8.85

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
51 - 60	99	73	70 - 77	69 - 77
50	98	71	68 - 75	67 - 75
49	98	70	67 - 73	66 - 74
48	97	68	65 - 71	64 - 72
47	95	67	64 - 70	63 - 71
46	94	65	62 - 68	61 - 69
45	92	64	61 - 67	60 - 68
44	90	63	60 - 66	59 - 67
43	88	62	58 - 65	58 - 66
42	85	61	57 - 64	57 - 65
41	83	60	56 - 63	56 - 64
40	80	59	55 - 62	55 - 63
39	77	57	54 - 60	53 - 61
38	73	56	53 - 59	52 - 60
37	70	55	52 - 58	51 - 59
36	68	55	52 - 58	51 - 59
35	65	54	51 - 57	50 - 58
34	62	53	50 - 56	49 - 57
33	58	52	49 - 55	48 - 56
32	55	51	48 - 54	47 - 55
31	51	50	47 - 53	46 - 54
30	47	49	46 - 52	45 - 53
29	42	48	45 - 51	44 - 52
28	39	47	44 - 50	43 - 51
27	35	46	43 - 49	42 - 50
26	32	45	42 - 49	41 - 49
25	30	45	42 - 48	41 - 49
24	28	44	41 - 47	40 - 48
23	26	44	41 - 47	40 - 48
22	24	43	40 - 46	39 - 47
21	22	42	39 - 45	38 - 46
20	21	42	39 - 45	38 - 46
19	19	41	38 - 44	37 - 45
18	16	40	37 - 43	36 - 44
17	13	39	36 - 42	35 - 43
16	11	38	34 - 41	34 - 42
15	8	36	33 - 39	32 - 40
14	7	35	32 - 38	31 - 39
13	5	34	30 - 37	30 - 38
12	3	31	28 - 35	27 - 35
11	2	29	26 - 32	25 - 33
0 - 10	1	28	25 - 31	24 - 32

© 2003 – 2013 Team Focus Limited

Descriptions of norms for open tests

Verbal level 1

General population

Composition of norm group			
Mean age	31.06	Male / Female percentage	61.95/38.05
SD age	9.69	'White' / 'Non white' percentage	68.81/31.19
Size of norm group	2930	Reliability	0.78
Mean	16.07	SEM (raw scores)	4.30
SD	9.17	SED (68/80/95%; raw scores)	6.08/7.78/12.16

Year 10 to 12 students

Composition of norm group			
Mean age	16.48	Male / Female percentage	58.21/41.79
SD age	0.53	'White' / 'Non white' percentage	60.70/39.30
Size of norm group	263	Reliability	0.79
Mean	27.18	SEM (raw scores)	3.69
SD	8.05	SED (68/80/95%; raw scores)	5.22/6.68/10.43

Verbal level 2

Undergraduates

Composition of norm group			
Mean age	21.87	Male / Female percentage	49.93/50.07
SD age	2.41	'White' / 'Non white' percentage	57.4742.53
Size of norm group	17,223	Reliability	0.84
Mean	30.59	SEM (raw scores)	4.19
SD	10.48	SED (68/80/95%; raw scores)	5.93/7.59/11.85

Postgraduates

Composition of norm group			
Mean age	24.93	Male / Female percentage	53.77/46.23
SD age	3.77	'White' / 'Non white' percentage	53.20/46.80
Size of norm group	1203	Reliability	0.76
Mean	31.49	SEM (raw scores)	5.67
SD	11.57	SED (68/80/95%; raw scores)	8.01/10.26/16.02

© 2003 – 2013 Team Focus Limited

Numerical level 1

General population

Composition of norm group			
Mean age	30.56	Male / Female percentage	68.01/31.99
SD age	9.58	'White' / 'Non white' percentage	68.45/31.55
Size of norm group	1287	Reliability	0.88
Mean	25.95	SEM (raw scores)	2.49
SD	7.20	SED (68/80/95%; raw scores)	3.53/4.51/7.05

Year 10 to 12 students

Composition of norm group			
Mean age	16.67	Male / Female percentage	56.10/43.90
SD age	0.50	'White' / 'Non white' percentage	49.38/50.62
Size of norm group	356	Reliability	0.91
Mean	28.87	SEM (raw scores)	1.72
SD	5.76	SED (68/80/95%; raw scores)	2.44/3.13/4.89

Numerical level 2

Undergraduates

Composition of norm group			
Mean age	22.79	Male / Female percentage	53.75/46.25
SD age	3.18	'White' / 'Non white' percentage	58.27/41.73
Size of norm group	27,336	Reliability	0.88
Mean	17.73	SEM (raw scores)	2.29
SD	6.62	SED (68/80/95%; raw scores)	3.24/4.15/6.49

Postgraduates

Composition of norm group			
Mean age	25.10	Male / Female percentage	58.94/41.06
SD age	3.57	'White' / 'Non white' percentage	54.87/45.13
Size of norm group	2012	Reliability	0.92
Mean	19.82	SEM (raw scores)	1.71
SD	6.06	SED (68/80/95%; raw scores)	2.42/3.10/4.85

© 2003 – 2013 Team Focus Limited

Abstract level 1

General population

Composition of norm group			
Mean age	28.16	Male / Female percentage	58.43/41.57
SD age	9.39	'White' / 'Non white' percentage	63.61/36.39
Size of norm group	453	Reliability	0.90
Mean	45.27	SEM (raw scores)	4.75
SD	15.03	SED (68/80/95%; raw scores)	6.72/8.60/13.43

Year 10 to 12 students

Composition of norm group			
Mean age	16.65	Male / Female percentage	61.54/38.46
SD age	0.60	'White' / 'Non white' percentage	36.92/63.08
Size of norm group	105	Reliability	0.89
Mean	46.43	SEM (raw scores)	3.32
SD	10.00	SED (68/80/95%; raw scores)	4.69/6.00/9.38

Abstract level 2

Undergraduates

Composition of norm group			
Mean age	24.57	Male / Female percentage	54.97/45.03
SD age	4.96	'White' / 'Non white' percentage	62.34/37.66
Size of norm group	10,464	Reliability	0.87
Mean	33.54	SEM (raw scores)	4.22
SD	11.73	SED (68/80/95%; raw scores)	5.95/7.65/11.96

Postgraduates

Composition of norm group			
Mean age	24.9	Male / Female percentage	51.19/48.81
SD age	3.43	'White' / 'Non white' percentage	57.86/42.14
Size of norm group	733	Reliability	0.91
Mean	37.27	SEM (raw scores)	3.97
SD	13.22	SED (68/80/95%; raw scores)	5.61/7.1811.22

© 2003 – 2013 Team Focus Limited

Combined reasoning test

Combined Reasoning Test: Composition of norm group			
Mean age	15.6	Male / Female percentage	51.7 / 48.3
SD age	1.2	'White' / 'Non white' percentage	88.9 / 11.1
Size of norm group	613		
Verbal section			
Mean	14.0	SEM (raw scores)	1.9
SD	4.7	SED (68/80/95%; raw scores)	2.7 / 3.4 / 5.3
Reliability	0.84		
Numerical section			
Mean	12.4	SEM (raw scores)	1.6
SD	4.2	SED (68/80/95%; raw scores)	2.2 / 2.8 / 4.4
Reliability	0.86		
Abstract section			
Mean	17.0	SEM (raw scores)	2.3
SD	5.9	SED (68/80/95%; raw scores)	3.2 / 4.1 / 6.5
Reliability	0.85		

© 2003 – 2013 Team Focus Limited

Additional norms for closed tests - Supplement 1

Description of norm group: Graduate applicants to a leading financial services institution. Compiled 2005.

Test: Verbal Reasoning Level: 3

Composition of norm group			
Mean age	22.9	Male / Female percentage	66.9 / 32.4
SD age	2.3	'White' / 'Non white' percentage	65.2 / 34.8

Size of norm group	420	Reliability	0.81
Mean	27.7	SEM (raw scores)	2.40
SD	5.5	SED (68/80/95%; raw scores)	3.39 / 4.34 / 6.78

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
38 or above	99	77	75 - 80	74 - 80
37	99	72	70 - 74	69 - 75
36	96	68	66 - 70	65 - 71
35	92	64	62 - 67	61 - 67
34	88	61	59 - 64	58 - 65
33	82	59	57 - 62	56 - 62
32	76	57	55 - 59	54 - 60
31	70	55	53 - 57	52 - 58
30	63	53	51 - 56	50 - 56
29	56	51	49 - 54	48 - 55
28	49	50	47 - 52	47 - 53
27	43	48	46 - 51	45 - 51
26	36	46	44 - 49	43 - 50
25	29	45	42 - 47	42 - 48
24	23	43	40 - 45	40 - 46
23	18	41	38 - 43	38 - 44
22	14	39	37 - 41	36 - 42
21	10	37	35 - 40	34 - 40
20	7	36	33 - 38	32 - 39
19	6	34	32 - 36	31 - 37
18	4	33	30 - 35	30 - 36
17	3	31	29 - 34	28 - 34
16	2	30	27 - 32	27 - 33
15	2	29	26 - 31	25 - 32
14	1	28	25 - 30	25 - 31
13 or below	1	26	24 - 29	23 - 30

© 2003 – 2013 Team Focus Limited

Test: Numerical Reasoning Level: 3

Composition of norm group			
Mean age	22.9	Male / Female percentage	67.6 / 32.4
SD age	2.3	'White' / 'Non white' percentage	65.2 / 34.4

Size of norm group	435	Reliability	0.92
Mean	23.9	SEM (raw scores)	1.73
SD	6.1	SED (68/80/95%; raw scores)	2.44 / 3.12 / 4.88

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
36 or above	99	78	76 - 80	76 - 80
35	99	72	71 - 74	70 - 75
34	97	68	67 - 70	66 - 71
33	94	65	64 - 67	63 - 68
32	91	63	61 - 65	61 - 65
31	87	61	60 - 63	59 - 63
30	83	59	58 - 61	57 - 62
29	77	58	56 - 59	55 - 60
28	72	56	54 - 58	54 - 58
27	67	54	53 - 56	52 - 56
26	60	53	51 - 54	50 - 55
25	54	51	49 - 53	49 - 53
24	48	49	48 - 51	47 - 52
23	43	48	47 - 50	46 - 50
22	38	47	45 - 49	45 - 49
21	33	46	44 - 47	43 - 48
20	27	44	42 - 46	42 - 46
19	22	42	41 - 44	40 - 44
18	18	41	39 - 42	38 - 43
17	14	39	37 - 41	37 - 41
16	11	38	36 - 40	36 - 40
15	9	36	35 - 38	34 - 39
14	6	35	33 - 37	33 - 37
13	4	33	31 - 35	31 - 35
12	3	31	29 - 32	28 - 33
11	1	28	26 - 30	26 - 30
10 or below	<1	25	23 - 27	23 - 27

© 2003 – 2013 Team Focus Limited

Test: Abstract Reasoning Level: 3

Composition of norm group			
Mean age	22.9	Male / Female percentage	66.3 / 33.7
SD age	2.3	'White' / 'Non white' percentage	65.8 / 34.2

Size of norm group	415	Reliability	0.92
Mean	35.7	SEM (raw scores)	3.14
SD	11.1	SED (68/80/95%; raw scores)	4.44 / 5.68 / 8.88

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
56 or above	99	73	69 - 76	69 - 77
55	98	70	67 - 74	66 - 75
54	96	68	65 - 71	64 - 72
53	95	66	63 - 69	62 - 70
52	93	65	62 - 68	61 - 69
51	91	64	61 - 67	60 - 68
50	90	63	59 - 66	59 - 67
49	87	61	58 - 65	57 - 65
48	85	60	57 - 63	56 - 64
47	82	59	56 - 62	55 - 63
46	79	58	55 - 61	54 - 62
45	77	57	54 - 61	53 - 61
44	75	57	53 - 60	53 - 61
43	72	56	53 - 59	52 - 60
42	69	55	52 - 58	51 - 59
41	67	54	51 - 57	50 - 58
40	64	54	50 - 57	49 - 58
39	60	53	49 - 56	49 - 57
38	56	52	48 - 55	48 - 56
37	52	51	47 - 54	47 - 55
36	49	50	47 - 53	46 - 54
35	45	49	46 - 52	45 - 53
34	42	48	45 - 51	44 - 52
33	39	47	44 - 50	43 - 51
32	36	46	43 - 49	42 - 50
31	33	46	42 - 49	42 - 50
30	30	45	42 - 48	41 - 49
29	28	44	41 - 47	40 - 48
28	26	43	40 - 47	39 - 47
27	23	43	39 - 46	39 - 47

Table continued overleaf . .

© 2003 – 2013 Team Focus Limited

Abstract Reasoning Level 3 (cont)

26	21	42	39 - 45	38 - 46
25	18	41	38 - 44	37 - 45
24	16	40	37 - 43	36 - 44
23	14	39	36 - 42	35 - 43
22	12	38	35 - 42	34 - 43
21	12	38	35 - 41	34 - 42
20	11	37	34 - 41	33 - 42
19	9	37	34 - 40	33 - 41
18	8	36	33 - 39	32 - 40
17	7	35	32 - 38	31 - 39
16	6	34	31 - 37	30 - 38
15	4	33	30 - 36	29 - 37
14	3	31	28 - 34	27 - 35
13	2	30	27 - 33	26 - 34
12	2	29	26 - 32	25 - 33
11 or below	1	27	23 - 30	23 - 31

© 2003 – 2013 Team Focus Limited

Test: Verbal Reasoning Level: 4

Description of norm group: Postgraduate students at a business school of a UK London University. Compiled 2005.

Composition of norm group			
Mean age	25.1	Male / Female percentage	55.1 / 44.9
SD age	3.8	'White' / 'Non white' percentage	41.1 / 58.9

Size of norm group	894	Reliability	0.90
Mean	23.0	SEM (raw scores)	2.2
SD	6.9	SED (68/80/95%; raw scores)	3.1 / 3.9 / 6.2

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
36 or above	99	74	71 - 78	70 - 78
35	98	71	68 - 74	67 - 75
34	97	69	65 - 72	65 - 73
33	95	66	63 - 69	62 - 70
32	92	64	61 - 67	60 - 68
31	89	62	59 - 65	58 - 66
30	85	60	57 - 63	56 - 64
29	80	59	55 - 62	55 - 63
28	75	57	54 - 60	53 - 61
27	70	55	52 - 58	51 - 59
26	64	54	50 - 57	49 - 58
25	58	52	49 - 55	48 - 56
24	52	51	47 - 54	47 - 55
23	47	49	46 - 52	45 - 53
22	41	48	45 - 51	44 - 52
21	35	46	43 - 49	42 - 50
20	30	45	42 - 48	41 - 49
19	26	43	40 - 47	39 - 47
18	22	42	39 - 45	38 - 46
17	18	41	38 - 44	37 - 45
16	15	40	37 - 43	36 - 44
15	12	38	35 - 42	34 - 42
14	10	37	34 - 40	33 - 41
13	8	36	33 - 39	32 - 40
12	7	35	32 - 38	31 - 39
11	5	34	31 - 37	30 - 38
10	4	33	30 - 36	29 - 37
9	4	32	29 - 35	28 - 36
8	3	32	29 - 35	28 - 36
7	3	31	28 - 34	27 - 35
6	3	31	28 - 34	27 - 35
5	3	30	27 - 34	26 - 34
4	2	29	26 - 32	25 - 33
3 or below	1	26	23 - 29	22 - 30

© 2003 – 2013 Team Focus Limited

Test: Numerical Reasoning Level: 4

Composition of norm group			
Mean age	24.7	Male / Female percentage	58.4 / 41.6
SD age	3.6	'White' / 'Non white' percentage	46.6 / 53.4

Size of norm group	1211	Reliability	0.89
Mean	17.6	SEM (raw scores)	2.2
SD	6.5	SED (68/80/95%; raw scores)	3.0 / 3.9 / 6.1

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
35 or above	99	77	74 - 81	73 - 82
34	99	74	71 - 78	70 - 79
33	99	72	69 - 76	68 - 77
32	98	70	67 - 74	66 - 75
31	97	69	65 - 72	64 - 73
30	96	67	64 - 71	63 - 72
29	95	66	63 - 70	62 - 71
28	93	65	62 - 68	61 - 69
27	91	63	60 - 66	59 - 67
26	88	62	58 - 65	57 - 66
25	85	60	57 - 64	56 - 64
24	82	59	56 - 62	55 - 63
23	79	58	55 - 61	54 - 62
22	75	57	53 - 60	52 - 61
21	71	56	52 - 59	51 - 60
20	67	54	51 - 58	50 - 59
19	62	53	50 - 56	49 - 57
18	57	52	49 - 55	48 - 56
17	51	50	47 - 54	46 - 54
16	44	49	45 - 52	44 - 53
15	38	47	44 - 50	43 - 51
14	32	45	42 - 49	41 - 50
13	27	44	40 - 47	39 - 48
12	21	42	38 - 45	38 - 46
11	15	40	36 - 43	35 - 44
10	11	38	34 - 41	33 - 42
9	8	36	33 - 39	32 - 40
8	5	34	31 - 37	30 - 38
7	3	32	28 - 35	27 - 36
6	2	29	25 - 32	25 - 33
5 or below	1	26	23 - 29	22 - 30

© 2003 – 2013 Team Focus Limited

Test: Abstract Reasoning Level: 4

Composition of norm group			
Mean age	25.4	Male / Female percentage	53.2 / 46.8
SD age	4.9	'White' / 'Non white' percentage	51.2 / 48.8

Size of norm group	573	Reliability	0.90
Mean	28.9	SEM (raw scores)	3.3
SD	10.3	SED (68/80/95%; raw scores)	4.6 / 5.9 / 9.2

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
51 or above	99	75	72 - 78	71 - 79
50	99	73	70 - 76	69 - 77
49	98	72	69 - 75	68 - 76
48	98	70	67 - 73	66 - 74
47	97	68	65 - 72	64 - 72
46	95	67	63 - 70	63 - 71
45	94	65	62 - 69	61 - 69
44	93	64	61 - 68	60 - 68
43	91	63	60 - 66	59 - 67
42	89	62	59 - 65	58 - 66
41	87	61	58 - 64	57 - 65
40	85	60	57 - 64	56 - 64
39	83	59	56 - 63	55 - 63
38	80	58	55 - 61	54 - 62
37	77	57	54 - 60	53 - 61
36	73	56	53 - 59	52 - 60
35	70	55	52 - 58	51 - 59
34	67	54	51 - 57	50 - 58
33	63	53	50 - 57	49 - 57
32	60	53	49 - 56	49 - 57
31	57	52	48 - 55	48 - 56
30	53	51	48 - 54	47 - 55
29	49	50	47 - 53	46 - 54
28	46	49	46 - 52	45 - 53
27	42	48	45 - 51	44 - 52
26	39	47	44 - 50	43 - 51
25	35	46	43 - 49	42 - 50

Table continued overleaf . . .

© 2003 – 2013 Team Focus Limited

Abstract Reasoning Level 4 (Cont)

24	32	45	42 - 49	41 - 49
23	29	45	41 - 48	40 - 49
22	26	44	40 - 47	40 - 48
21	24	43	40 - 46	39 - 47
20	21	42	39 - 45	38 - 46
19	20	41	38 - 45	37 - 46
18	18	41	38 - 44	37 - 45
17	16	40	37 - 43	36 - 44
16	13	39	36 - 42	35 - 43
15	11	38	35 - 41	34 - 42
14	9	37	34 - 40	33 - 41
13	7	35	32 - 39	31 - 40
12	5	34	30 - 37	30 - 38
11	3	31	28 - 34	27 - 35
10	2	29	26 - 32	25 - 33
9	1	27	24 - 30	23 - 31
8 or below	1	25	22 - 28	21 - 29

© 2003 – 2013 Team Focus Limited

Test: Verbal Reasoning Level: 4

Description of norm group: British MENSA members. Compiled 2005.

Composition of norm group			
Mean age	43.4	Male / Female percentage	49.7 / 50.3
SD age	12.2	'White' / 'Non white' percentage	95.0 / 5.0

Size of norm group	193	Reliability	0.75
Mean	29.3	SEM (raw scores)	2.3
SD	4.6	SED (68/80/95%; raw scores)	3.25 / 4.16 / 6.5

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
38 or above	99	81	79 - 84	78 - 84
37	99	72	70 - 75	69 - 75
36	96	67	65 - 70	64 - 70
35	91	64	61 - 66	61 - 67
34	85	60	58 - 63	57 - 63
33	78	58	55 - 60	55 - 61
32	69	55	53 - 57	52 - 58
31	60	53	50 - 55	50 - 56
30	51	50	48 - 52	47 - 53
29	42	48	46 - 50	45 - 51
28	35	46	44 - 48	43 - 49
27	27	44	42 - 46	41 - 47
26	21	42	40 - 44	39 - 45
25	17	41	38 - 43	38 - 44
24	14	39	37 - 42	36 - 42
23	11	38	35 - 40	35 - 40
22	8	36	33 - 38	33 - 39
21	6	34	32 - 36	31 - 37
20	4	33	30 - 35	30 - 35
19	3	31	28 - 33	28 - 34
18	2	29	26 - 31	26 - 32
17 or below	1	27	25 - 29	24 - 30

© 2003 – 2013 Team Focus Limited

Test: Numerical Reasoning Level: 4

Composition of norm group			
Mean age	41.9	Male / Female percentage	60.5 / 39.5
SD age	12.0	'White' / 'Non white' percentage	93.9 / 6.1

Size of norm group	220	Reliability	0.90
Mean	19.4	SEM (raw scores)	1.96
SD	6.2	SED (68/80/95%; raw scores)	2.77 / 3.55 / 5.54

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
36	99	78	76 - 80	75 - 80
35	99	73	71 - 75	71 - 76
34	98	70	68 - 72	67 - 72
33	96	68	66 - 70	65 - 70
32	95	66	64 - 68	64 - 69
31	94	65	63 - 67	63 - 68
30	92	64	62 - 66	62 - 67
29	91	63	61 - 65	61 - 66
28	88	62	60 - 64	59 - 64
27	86	61	59 - 63	58 - 63
26	83	60	58 - 62	57 - 62
25	82	59	57 - 61	57 - 62
24	80	58	56 - 60	56 - 61
23	77	57	55 - 59	55 - 60
22	72	56	54 - 58	53 - 58
21	68	55	53 - 57	52 - 57
20	62	53	51 - 55	51 - 56
19	55	51	49 - 53	49 - 54
18	47	49	47 - 51	47 - 52
17	38	47	45 - 49	44 - 49
16	30	45	43 - 47	42 - 47
15	24	43	41 - 45	40 - 45
14	18	41	39 - 43	38 - 43
13	13	39	37 - 41	36 - 41
12	10	37	35 - 39	35 - 40
11	7	35	33 - 37	33 - 38
10	4	32	30 - 34	30 - 35
9	2	29	27 - 31	26 - 31
8 or below	1	25	23 - 27	23 - 28

© 2003 – 2013 Team Focus Limited

Test: Abstract Reasoning Level: 4

Composition of norm group			
Mean age	43.0	Male / Female percentage	56.3 / 43.7
SD age	12.8	'White' / 'Non white' percentage	94.8 / 5.2

Size of norm group	144	Reliability	0.91
Mean	35.1	SEM (raw scores)	3.12
SD	10.4	SED (68/80/95%; raw scores)	4.41 / 5.65 / 8.82

Raw score	Percentile rank	T-score	68% T-score confidence band	80% T-score confidence band
53 or above	99	76	73 - 79	72 - 80
52	98	71	68 - 74	67 - 75
51	97	68	65 - 71	64 - 72
50	94	66	63 - 69	62 - 70
49	92	64	61 - 67	60 - 68
48	89	62	59 - 65	58 - 66
47	86	61	58 - 64	57 - 65
46	82	59	56 - 62	55 - 63
45	78	58	55 - 61	54 - 62
44	74	56	53 - 59	52 - 60
43	70	55	52 - 58	51 - 59
42	68	55	51 - 58	51 - 59
41	66	54	51 - 57	50 - 58
40	63	53	50 - 56	49 - 57
39	59	52	49 - 56	48 - 56
38	57	52	49 - 55	48 - 56
37	55	51	48 - 54	47 - 55
36	52	51	47 - 54	47 - 55
35	49	50	47 - 53	46 - 54
34	47	49	46 - 52	45 - 53
33	45	49	46 - 52	45 - 53
32	41	48	45 - 51	44 - 52
31	37	47	43 - 50	43 - 51
30	32	45	42 - 48	41 - 49
29	28	44	41 - 47	40 - 48
28	24	43	40 - 46	39 - 47
27	20	42	39 - 45	38 - 46
26	18	41	38 - 44	37 - 45
25	17	40	37 - 43	36 - 44
24	15	40	37 - 43	36 - 44
23	14	39	36 - 42	35 - 43
22	12	38	35 - 42	34 - 42
21	11	38	34 - 41	34 - 42
20	9	36	33 - 40	32 - 40
19	7	35	32 - 38	31 - 39
18	5	34	31 - 37	30 - 38
17	4	32	29 - 35	28 - 36
16	2	29	26 - 32	25 - 33
15 or below	1	26	23 - 30	22 - 30

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Descriptions of additional norms for open tests - Supplement 1

Internal applicants from a public service organisation completing the test as preparation for an internal selection process. Compiled 2007.

Verbal level 1: Composition of norm group			
Mean age	35.2	Male / Female percentage	74.3/25.7
SD age	8.4	'White' / 'Non white' percentage	85.0/15.0
Size of norm group	768	Reliability	0.91
Mean	23.8	SEM (raw scores)	2.3
SD	7.6	SED (68/80/95%; raw scores)	3.2/4.1/6.4

Numerical level 1: Composition of norm group			
Mean age	35.7	Male / Female percentage	91.3 / 8.7
SD age	8.2	'White' / 'Non white' percentage	86.8 / 13.2
Size of norm group	689	Reliability	0.86
Mean	23.2	SEM (raw scores)	1.7
SD	4.6	SED (68/80/95%; raw scores)	2.4 / 3.1 / 4.9

Year 10 to 12 students in compulsory education from non-selective schools

Combined Reasoning Test: Composition of norm group			
Mean age	15.7	Male / Female percentage	63.2 / 36.8
SD age	2.4	'White' / 'Non white' percentage	86.7 / 13.3
Size of norm group	345		
Verbal section			
Mean	12.6	SEM (raw scores)	2.0
SD	4.7	SED (68/80/95%; raw scores)	2.9/3.7/5.8
Reliability	0.81		
Numerical section			
Mean	11.6	SEM (raw scores)	2.1
SD	4.5	SED (68/80/95%; raw scores)	2.9/3.7/5.8
Reliability	0.79		
Abstract section			
Mean	14.9	SEM (raw scores)	2.7
SD	6.3	SED (68/80/95%; raw scores)	3.8/4.87.6
Reliability	0.82		

© 2003 – 2013 Team Focus Limited

Year 10 to 12 students in compulsory education from selective schools

Combined Reasoning Test: Composition of norm group			
Mean age	15.2	Male / Female percentage	37.6 / 62.4
SD age	1.0	'White' / 'Non white' percentage	89.3 / 10.7
Size of norm group	290		
Verbal section			
Mean	15.5	SEM (raw scores)	1.3
SD	4.2	SED (68/80/95%; raw scores)	1.9/2.4/3.8
Reliability	0.90		
Numerical section			
Mean	14.0	SEM (raw scores)	1.2
SD	2.8	SED (68/80/95%; raw scores)	1.7/2.2/3.4
Reliability	0.82		
Abstract section			
Mean	19.4	SEM (raw scores)	2.5
SD	4.9	SED (68/80/95%; raw scores)	3.5/4.5/7.1
Reliability	0.74		

© 2003 – 2013 Team Focus Limited

Graduate applicants for Financial Services (Investments): Verbal Reasoning Level 2

Norms are presented for Level 2 of the Verbal Reasoning Test ('closed' version). These norms were derived from a sample of applicants for positions in 'Operations' at a multinational investment and fund management company. The data from this sample was collected between 2006 and 2012.

The sample characteristics and overall summary statistics are presented firstly, followed by the norm tables themselves.

Sample characteristics and ethnic composition

Sample size	Mean age	% male	% female
1942	30.05	42	58

	Ethnic Group	Percent
1	White British	54.07
2	White Irish	1.60
3	Other white	13.54
4	White + Black Caribbean	0.93
5	White + Black African	0.36
6	White Asian	0.88
7	Other mixed	1.60
8	Indian	8.96
9	Pakistani	1.75
10	Bangladeshi	1.13
11	Other Asian	1.54
12	Caribbean	3.09
13	African	5.82
14	Other Black	0.72
15	Chinese	2.06
16	Any other	0.41
99	Not indicated	1.54

Means and standard deviations

	Mean	SD
Raw score	18.94	5.33
Prop. Correct of attempted	0.70	0.14
N Attempted	27.12	5.12

© 2003 – 2013 Team Focus Limited

Raw score Norms: Verbal Reasoning Level 2

Raw score	Percentile	T Score	z Score
0	1	2	-4.75
1	1	12	-3.75
2	1	17	-3.32
3	1	19	-3.12
4	1	21	-2.87
5	1	24	-2.62
6	1	26	-2.41
7	1	28	-2.18
8	3	31	-1.93
9	4	33	-1.72
10	6	35	-1.54
11	9	36	-1.37
12	11	38	-1.20
13	15	40	-1.03
14	19	41	-0.87
15	24	43	-0.71
16	29	45	-0.55
17	35	46	-0.39
18	41	48	-0.23
19	47	49	-0.07
20	54	51	0.09
21	61	53	0.27
22	68	55	0.47
23	75	57	0.68
24	82	59	0.90
25	87	61	1.12
26	91	64	1.37
27	95	67	1.66
28	98	70	2.00
29	99	74	2.39
30	99	79	2.89
31	99	86	3.60
32	99	86	3.60

© 2003 – 2013 Team Focus Limited

Graduate applicants for Financial Services (Insurance): Level 3 Reasoning Tests

Norms are presented for Level 3 of each of the Numerical, Verbal and Abstract Reasoning Tests ('closed' versions). These norms were derived from a sample of applicants for graduate positions at a multinational insurance company. The data from these samples was collected between January and May 2005.

The sample characteristics and overall summary statistics are presented firstly, followed by the norm tables themselves.

Sample characteristics and ethnic composition

Test	Sample size	Mean age	% male	% female
Numerical	483	22.89	67%	33%
Verbal	465	22.86	66%	28%
Abstract	462	22.90	66%	34%

		Percentage		
	Ethnic Group	**Numerical**	**Verbal**	**Abstract**
1	White British	48.24	48.39	48.05
2	White Irish	3.52	3.66	3.68
3	Other white	14.08	13.98	14.50
4	White + Black Caribbean	0.41	0.43	0.43
5	White + Black African	0.41	0.43	0.43
6	White Asian	1.45	1.51	1.52
7	Other mixed	1.04	1.08	1.30
8	Indian	15.11	14.62	14.50
9	Pakistani	2.69	2.80	2.60
10	Bangladeshi	1.45	1.51	1.52
11	Other Asian	1.86	2.15	1.95
12	Caribbean	0.00	0.00	0.00
13	African	3.31	2.80	2.81
14	Other Black	4.97	0.00	0.00
15	Chinese	1.24	5.16	5.19
16	Any other	0.21	1.29	1.52
99	Not indicated	0.00	0.22	0.00

© 2003 – 2013 Team Focus Limited

Means and standard deviations

		Mean	SD
Numerical	Raw score	23.872	6.105
	Prop. Correct of attempted	0.875	0.106
	N Attempted	27.273	6.108
Verbal	Raw score	27.798	5.351
	Prop. Correct of attempted	0.766	0.105
	N Attempted	36.297	4.809
Abstract	Raw score	35.831	11.104
	Prop. Correct of attempted	0.692	0.174
	N Attempted	51.831	9.741

© 2003 – 2013 Team Focus Limited

Raw score Norms: Numerical Reasoning Level 3

Raw score	Percentile	T Score	z Score
0	1	2	-4.75
1	1	2	-4.75
2	1	2	-4.75
3	1	14	-3.57
4	1	19	-3.08
5	1	21	-2.91
6	1	22	-2.84
7	1	23	-2.73
8	1	24	-2.57
9	1	25	-2.47
10	1	27	-2.33
11	2	29	-2.12
12	3	31	-1.91
13	4	33	-1.70
14	7	35	-1.51
15	9	36	-1.35
16	11	38	-1.21
17	14	39	-1.08
18	18	41	-0.93
19	22	42	-0.77
20	27	44	-0.60
21	33	46	-0.43
22	39	47	-0.29
23	43	48	-0.17
24	48	50	-0.05
25	54	51	0.10
26	61	53	0.28
27	67	54	0.44
28	73	56	0.61
29	78	58	0.77
30	83	60	0.96
31	87	61	1.14
32	91	63	1.33
33	94	66	1.55
34	97	68	1.83
35	99	72	2.19
36	99	77	2.73

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Raw score Norms: Verbal Reasoning Level 3

Raw score	Percentile	T Score	z Score
0	1	2	-4.75
1	1	2	-4.75
2	1	2	-4.75
3	1	2	-4.75
4	1	2	-4.75
5	1	2	-4.75
6	1	2	-4.75
7	1	14	-3.56
8	1	19	-3.07
9	1	21	-2.88
10	1	21	-2.87
11	1	21	-2.85
12	1	24	-2.59
13	1	27	-2.30
14	2	29	-2.13
15	2	30	-2.04
16	3	31	-1.94
17	4	32	-1.80
18	5	33	-1.67
19	6	34	-1.55
20	8	36	-1.42
21	10	37	-1.27
22	14	39	-1.10
23	18	41	-0.92
24	23	43	-0.73
25	29	45	-0.55
26	36	46	-0.37
27	42	48	-0.20
28	49	50	-0.03
29	56	51	0.14
30	62	53	0.32
31	69	55	0.50
32	76	57	0.69
33	82	59	0.90
34	87	61	1.14
35	93	64	1.44

© 2003 – 2013 Team Focus Limited

36	97	68	1.84
37	99	73	2.28
38	99	79	2.87
39	99	79	2.87
40	99	79	2.87

© 2003 – 2013 Team Focus Limited

Raw score Norms: Abstract Reasoning Level 3

Raw score	Percentile	T Score	z Score
0	1	2	-4.75
1	1	2	-4.75
2	1	2	-4.75
3	1	2	-4.75
4	1	2	-4.75
5	1	2	-4.75
6	1	2	-4.75
7	1	2	-4.75
8	1	2	-4.75
9	1	2	-4.75
10	1	22	-2.79
11	1	26	-2.38
12	2	29	-2.15
13	2	30	-2.01
14	3	31	-1.88
15	4	33	-1.73
16	6	34	-1.60
17	7	35	-1.51
18	8	36	-1.43
19	9	36	-1.35
20	10	37	-1.27
21	11	38	-1.21
22	12	38	-1.16
23	14	39	-1.08
24	16	40	-0.99
25	18	41	-0.90
26	21	42	-0.82
27	23	43	-0.74
28	25	43	-0.67
29	27	44	-0.61
30	29	45	-0.54
31	32	45	-0.47
32	35	46	-0.38
33	38	47	-0.30
34	42	48	-0.21
35	45	49	-0.13
36	48	50	-0.04

© 2003 – 2013 Team Focus Limited

37	52	50	0.04
38	55	51	0.14
39	59	52	0.23
40	63	53	0.32
41	66	54	0.41
42	68	55	0.48
43	71	56	0.55
44	74	56	0.64
45	76	57	0.72
46	79	58	0.80
47	82	59	0.90
48	85	60	1.02
49	87	61	1.15
50	90	63	1.27
51	92	64	1.38
52	93	65	1.49
53	95	66	1.62
54	97	68	1.82
55	98	71	2.07
56	99	73	2.29
57	99	75	2.52
58	99	79	2.88
59	99	79	2.88
60	99	79	2.88

© 2003 – 2013 Team Focus Limited

Graduate applicants for Financial Services (Investments): Numerical Reasoning Level 3

Norms are presented for Level 3 of the Numerical Reasoning Test ('closed' version). These norms were derived from a sample of applicants for graduate positions at a multinational investment and fund management company. The data from this sample was collected between 2006 and 2012.

The sample characteristics and overall summary statistics are presented firstly, followed by the norm tables themselves.

Sample characteristics and ethnic composition

Sample size	Mean age	% male	% female
2390	24.67	71	29

	Ethnic Group	Percent
1	White British	25.10
2	White Irish	2.89
3	Other white	24.39
4	White + Black Caribbean	0.29
5	White + Black African	0.63
6	White Asian	1.34
7	Other mixed	1.55
8	Indian	15.48
9	Pakistani	2.93
10	Bangladeshi	1.13
11	Other Asian	3.43
12	Caribbean	0.54
13	African	5.56
14	Other Black	0.13
15	Chinese	11.21
16	Any other	1.59
99	Not indicated	1.80

© 2003 – 2013 Team Focus Limited

Means and standard deviations

	Mean	SD
Raw score	25.25	5.88
Prop. Correct of attempted	0.88	0.11
N Attempted	28.80	5.76

© 2003 – 2013 Team Focus Limited

Raw score Norms: Numerical Reasoning Level 3

Raw score	Percentile	T Score	z Score
0	1	17	-3.32
1	1	20	-3.01
2	1	21	-2.92
3	1	21	-2.87
4	1	22	-2.77
5	1	23	-2.66
6	1	24	-2.58
7	1	25	-2.52
8	1	25	-2.49
9	1	25	-2.46
10	1	26	-2.38
11	1	27	-2.27
12	2	28	-2.15
13	2	30	-2.00
14	3	32	-1.82
15	5	34	-1.65
16	7	35	-1.49
17	9	37	-1.34
18	12	38	-1.18
19	15	40	-1.02
20	19	41	-0.87
21	24	43	-0.71
22	29	44	-0.56
23	34	46	-0.41
24	39	47	-0.27
25	45	49	-0.14
26	50	50	0.01
27	57	52	0.18
28	64	54	0.35
29	71	55	0.54
30	77	57	0.73
31	82	59	0.93
32	87	61	1.13
33	92	64	1.38
34	96	67	1.71
35	98	72	2.15
36	99	77	2.73

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Applicants to a leading UK Business School: Level 4 Reasoning Tests

Norms are presented for Level 4 of each of the Numerical, Verbal and Abstract Reasoning Tests ('closed' versions). These norms were derived from a sample of applicants to a leading UK business school and the sample consisted largely of individuals with degree level and / or professional qualifications. The data from these samples was collected between September 2003 and November 2010.

The sample characteristics and overall summary statistics are presented firstly, followed by the norm tables themselves.

Sample characteristics and ethnic composition

Test	Sample size	Mean age	% male	% female
Numerical	1408	24.88	61%	39%
Verbal	995	25.18	55%	41%
Abstract	646	25.46	58%	42%

		Percentage		
	Ethnic Group	**Numerical**	**Verbal**	**Abstract**
1	White British	10.23	8.80	8.51
2	White Irish	0.57	0.84	0.62
3	Other white	33.10	29.53	37.31
4	White + Black Caribbean	0.07	0.21	0.15
5	White + Black African	0.28	0.42	0.31
6	White Asian	1.21	0.52	0.77
7	Other mixed	0.92	0.84	1.08
8	Indian	12.50	13.30	10.53
9	Pakistani	0.99	0.63	0.77
10	Bangladeshi	0.57	0.31	0.15
11	Other Asian	7.10	5.65	5.42
12	Caribbean	0.64	0.63	0.15
13	African	4.90	5.76	5.42
14	Other Black	0.36	0.21	0.15
15	Chinese	20.67	26.60	21.36
16	Any other	1.56	1.57	1.70
99	Not indicated	4.33	4.19	5.57

© 2003 – 2013 Team Focus Limited

Means and standard deviations

		Mean	SD
Numerical	Raw score	17.63	6.57
	Prop. Correct of attempted	.80	.15
	N Attempted	22.20	7.31
Verbal	Raw score	23.11	6.62
	Prop. Correct of attempted	.73	.12
	N Attempted	31.91	7.78
Abstract	Raw score	28.87	10.14
	Prop. Correct of attempted	.57	.18
	N Attempted	51.27	10.35

© 2003 – 2013 Team Focus Limited

Raw score norms: Numerical Reasoning Level 4

Raw score	Percentile	T Score	z Score
0	1	16	-3.39
1	1	20	-3.00
2	1	22	-2.79
3	1	24	-2.62
4	1	26	-2.45
5	1	27	-2.26
6	2	29	-2.05
7	3	32	-1.81
8	6	34	-1.59
9	8	36	-1.40
10	11	38	-1.24
11	15	40	-1.04
12	20	42	-0.83
13	26	44	-0.64
14	32	45	-0.47
15	38	47	-0.31
16	44	49	-0.15
17	51	50	0.02
18	57	52	0.18
19	62	53	0.31
20	67	54	0.44
21	71	56	0.56
22	75	57	0.68
23	78	58	0.79
24	82	59	0.90
25	85	60	1.02
26	88	62	1.15
27	90	63	1.30
28	93	65	1.45
29	94	66	1.58
30	95	67	1.69
31	97	68	1.81
32	98	70	2.00
33	99	72	2.24
34	99	75	2.54
35	99	80	2.96
36	99	80	2.96

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Raw score norms: Verbal Reasoning Level 4

Raw score	Percentile	T Score	z Score
0	1	16	-3.38
1	1	19	-3.08
2	1	22	-2.84
3	1	24	-2.59
4	1	26	-2.38
5	1	27	-2.26
6	1	28	-2.18
7	2	29	-2.10
8	2	30	-2.00
9	3	31	-1.92
10	3	32	-1.82
11	4	33	-1.71
12	6	34	-1.57
13	8	36	-1.44
14	10	37	-1.31
15	12	38	-1.19
16	14	39	-1.06
17	18	41	-0.93
18	21	42	-0.81
19	25	43	-0.67
20	30	45	-0.54
21	35	46	-0.39
22	41	48	-0.22
23	47	49	-0.07
24	53	51	0.07
25	58	52	0.20
26	64	54	0.36
27	70	55	0.53
28	75	57	0.68
29	80	58	0.84
30	85	60	1.02
31	89	62	1.22
32	92	64	1.43
33	95	66	1.64
34	97	69	1.87
35	98	72	2.16
36	99	75	2.50

© 2003 – 2013 Team Focus Limited

37	99	78	2.82
38	99	82	3.20
39	99	82	3.20
40	99	82	3.20

© 2003 – 2013 Team Focus Limited

Raw score norms: Abstract Reasoning Level 4

Raw score	Percentile	T Score	z Score
0	1	16	-3.37
1	1	19	-3.08
2	1	20	-3.05
3	1	20	-3.02
4	1	21	-2.89
5	1	22	-2.81
6	1	22	-2.78
7	1	23	-2.68
8	1	25	-2.49
9	1	27	-2.29
10	2	29	-2.08
11	3	31	-1.86
12	5	33	-1.66
13	7	35	-1.50
14	8	36	-1.38
15	10	37	-1.28
16	12	38	-1.18
17	14	39	-1.07
18	17	40	-0.97
19	19	41	-0.89
20	21	42	-0.82
21	23	43	-0.73
22	26	44	-0.63
23	30	45	-0.53
24	33	46	-0.44
25	36	46	-0.35
26	40	47	-0.27
27	43	48	-0.17
28	47	49	-0.08
29	51	50	0.02
30	54	51	0.11
31	58	52	0.20
32	61	53	0.29
33	64	54	0.37
34	67	54	0.45
35	71	55	0.54

© 2003 – 2013 Team Focus Limited

36	74	56	0.64
37	77	57	0.75
38	80	59	0.85
39	83	60	0.95
40	85	61	1.05
41	87	61	1.13
42	89	62	1.21
43	91	63	1.33
44	93	65	1.47
45	94	66	1.59
46	96	67	1.70
47	97	68	1.84
48	98	70	1.99
49	98	71	2.14
50	99	73	2.30
51	99	74	2.42
52	99	76	2.59
53	99	79	2.95
54	99	79	2.95
55	99	79	2.95
56	99	79	2.95
57	99	79	2.95
58	99	79	2.95
59	99	79	2.95
60	99	79	2.95

© 2003 – 2013 Team Focus Limited

Senior professionals and administrators applying to a regulatory body: Level 4 Reasoning Tests

Norms are presented for Level 4 of each of the Numerical, Verbal and Abstract Reasoning Tests ('closed' versions). These norms were derived from a sample of applicants for positions on the panel of a regulatory body in the UK and the sample consisted largely of highly experienced individuals from a variety of professional and organisational roles with a mean age of approximately 53 years. The data from these samples was collected between July and September 2011. The sample characteristics and overall summary statistics are presented firstly, followed by the norm tables themselves.

Sample characteristics and ethnic composition

Test	Sample size	Mean age	% male	% female
Numerical	589	53.11	53%	47%
Verbal	595	53.08	53%	47%
Abstract	588	52.98	53%	47%

		Percentage		
	Ethnic Group	**Numerical**	**Verbal**	**Abstract**
1	White British	78.27	78.49	77.72
2	White Irish	4.92	4.87	4.93
3	Other white	3.57	3.53	3.40
4	White + Black Caribbean	0.17	0.17	0.17
5	White + Black African	0.68	0.67	0.68
6	White Asian	0.34	0.34	0.34
7	Other mixed	0.51	0.50	0.51
8	Indian	5.77	5.71	6.12
9	Pakistani	1.87	1.85	1.87
10	Bangladeshi	0.34	0.34	0.34
11	Other Asian	0.68	0.67	1.02
12	Caribbean	0.85	0.84	0.85
13	African	0.51	0.50	0.51
14	Other Black	0.17	0.17	0.17
15	Chinese	0.34	0.34	0.34
16	Any other	0.00	0.00	0.00
99	Not indicated	1.02	1.01	1.02

© 2003 – 2013 Team Focus Limited

Means and standard deviations

		Mean	SD
Numerical	Raw score	17.50	5.64
	Prop. Correct of attempted	0.77	0.16
	N Attempted	22.79	6.01
Verbal	Raw score	27.07	5.33
	Prop. Correct of attempted	0.78	0.10
	N Attempted	34.25	5.63
Abstract	Raw score	27.57	10.31
	Prop. Correct of attempted	0.60	0.16
	N Attempted	46.08	11.78

© 2003 – 2013 Team Focus Limited

Raw score norms: Numerical Reasoning Level 4

Raw score	Percentile	T Score	z Score
0	1	2	-4.75
1	1	2	-4.75
2	1	2	-4.75
3	1	2	-4.75
4	1	19	-3.05
5	1	24	-2.58
6	1	27	-2.26
7	2	30	-2.02
8	3	32	-1.81
9	6	34	-1.60
10	8	36	-1.39
11	12	38	-1.19
12	16	40	-1.00
13	21	42	-0.79
14	28	44	-0.58
15	35	46	-0.38
16	43	48	-0.17
17	51	50	0.02
18	57	52	0.18
19	63	53	0.32
20	68	55	0.47
21	74	56	0.64
22	79	58	0.80
23	83	59	0.94
24	86	61	1.07
25	89	62	1.23
26	92	64	1.40
27	94	66	1.56
28	96	67	1.72
29	97	69	1.91
30	98	71	2.09
31	99	72	2.25
32	99	74	2.44
33	99	76	2.64
34	99	79	2.87
35	99	82	3.22
36	99	82	3.22

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

Raw score norms: Verbal Reasoning Level 4

Raw score	Percentile	T Score	z Score
0	1	2	-4.75
1	1	2	-4.75
2	1	2	-4.75
3	1	2	-4.75
4	1	14	-3.63
5	1	19	-3.14
6	1	20	-2.96
7	1	21	-2.92
8	1	21	-2.93
9	1	21	-2.93
10	1	20	-2.98
11	1	21	-2.88
12	1	23	-2.66
13	1	26	-2.37
14	2	29	-2.12
15	2	30	-1.96
16	3	31	-1.85
17	4	33	-1.74
18	6	34	-1.58
19	8	36	-1.41
20	10	37	-1.27
21	13	39	-1.11
22	17	41	-0.94
23	22	42	-0.77
24	27	44	-0.61
25	33	46	-0.44
26	40	47	-0.27
27	46	49	-0.09
28	53	51	0.08
29	60	53	0.26
30	68	55	0.45
31	75	57	0.67
32	81	59	0.89
33	87	61	1.12
34	91	64	1.36
35	95	66	1.64
36	98	70	2.00

© 2003 – 2013 Team Focus Limited

37	99	75	2.51
38	99	83	3.35
39	99	83	3.35
40	99	83	3.35

© 2003 – 2013 Team Focus Limited

Raw score norms: Abstract Reasoning Level 4

Raw score	Percentile	T Score	z Score
0	1	2	-4.75
1	1	2	-4.75
2	1	2	-4.75
3	1	2	-4.75
4	1	14	-3.62
5	1	19	-3.14
6	1	17	-3.27
7	1	23	-2.72
8	1	28	-2.19
9	3	31	-1.87
10	4	33	-1.71
11	5	34	-1.61
12	7	35	-1.47
13	9	37	-1.32
14	12	38	-1.18
15	14	39	-1.08
16	16	40	-0.98
17	18	41	-0.90
18	21	42	-0.82
19	23	43	-0.75
20	25	43	-0.66
21	29	44	-0.56
22	32	45	-0.46
23	36	46	-0.37
24	39	47	-0.29
25	42	48	-0.21
26	44	49	-0.14
27	47	49	-0.07
28	50	50	0.00
29	53	51	0.08
30	57	52	0.18
31	61	53	0.28
32	64	54	0.37
33	67	54	0.44
34	70	55	0.52
35	73	56	0.62

© 2003 – 2013 Team Focus Limited

36	76	57	0.72
37	79	58	0.81
38	82	59	0.90
39	84	60	1.01
40	87	61	1.14
41	90	63	1.26
42	92	64	1.37
43	94	65	1.52
44	96	67	1.71
45	97	69	1.90
46	98	70	2.02
47	98	71	2.09
48	99	72	2.18
49	99	73	2.33
50	99	75	2.52
51	99	77	2.69
52	99	78	2.84
53	99	80	3.04
54	99	84	3.40
55	99	84	3.40
56	99	84	3.40
57	99	84	3.40
58	99	84	3.40
59	99	84	3.40
60	99	84	3.40

© 2003 – 2013 Team Focus Limited

Appendix 4: Comparison tables

The tables that follow allow you to compare the score a student achieves on the Verbal, Numerical and Abstract Reasoning tests with other norm groups, using a common scale. In each case take the relevant comparison (IRT) score from the Adviser's Report and look across the table to the relevant percentile for a particular comparison.

For example, if Student A achieves an IRT score of **100** (Verbal {Closed} Level-2), this puts him/her at the **61st** percentile when compared to students considering A-Level (or equivalent) qualifications. It **also** suggests that he/she is at the **40th** percentile when compared to undergraduate students.

VERBAL (CLOSED) standardised IRT score	**Percentiles**			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
40	1	1	1	1
41	1	1	1	1
42	1	1	1	1
43	1	1	1	1
44	1	1	1	1
45	2	1	1	1
46	2	1	1	1
47	2	1	1	1
48	2	1	1	1
49	3	1	1	1
50	3	1	1	1
51	4	1	1	1
52	5	1	1	1
53	6	1	1	1
54	7	1	1	1
55	8	1	1	1
56	10	1	1	1
57	11	1	1	1
58	12	1	1	1
59	14	1	1	1
60	16	1	1	1
61	18	1	1	1
62	20	1	1	1
63	22	1	2	1
64	23	1	2	1
65	24	1	2	1

© 2003 – 2013 Team Focus Limited

VERBAL (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
66	26	3	2	1
67	28	3	2	1
68	29	4	2	1
69	31	4	3	1
70	32	5	3	2
71	33	5	3	2
72	35	6	4	2
73	37	7	4	3
74	39	8	4	3
75	40	9	5	3
76	41	10	6	4
77	42	11	7	4
78	44	13	7	4
79	45	15	7	5
80	46	17	8	5
81	47	19	9	6
82	49	21	11	7
83	51	24	12	8
84	55	25	13	9
85	57	27	14	10
86	59	29	15	11
87	61	31	17	12
88	63	33	19	12
89	65	35	20	13
90	67	38	21	14
91	69	40	23	15
92	71	42	25	17
93	74	44	26	18
94	76	47	28	19
95	78	50	30	20
96	80	52	32	22
97	82	54	34	24
98	83	57	36	25
99	85	59	38	27
100	85	61	40	29
101	87	63	44	31
102	88	64	46	34

© 2003 – 2013 Team Focus Limited

VERBAL (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
103	89	65	48	36
104	91	67	50	38
105	91	68	52	39
106	92	69	55	41
107	93	71	58	42
108	94	73	60	44
109	94	75	62	45
110	94	77	65	47
111	95	79	67	49
112	95	81	69	50
113	96	84	71	51
114	96	85	74	53
115	96	86	77	54
116	96	87	79	57
117	96	89	81	60
118	97	89	83	63
119	97	90	85	65
120	97	91	86	67
121	97	92	88	69
122	97	93	89	70
123	97	94	90	73
124	98	95	91	75
125	98	96	92	77
126	98	96	93	79
127	98	96	94	80
128	98	97	95	82
129	98	98	95	84
130	98	98	96	85
131	98	98	96	86
132	98	98	97	87
133	98	98	97	88
134	98	99	97	90
135	98	99	97	90
136	98	99	98	91
137	98	99	99	92
138	98	99	99	93
139	98	99	99	93

© 2003 – 2013 Team Focus Limited

VERBAL (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
140	98	99	99	94
141	99	99	99	94
142	99	99	99	95
143	99	99	99	96
144	99	99	99	96
145	99	99	99	96
146	99	99	99	96
147	99	99	99	97
148	99	99	99	97
149	99	99	99	98
150	99	99	99	98
151	99	99	99	98
152	99	99	99	98
153	99	99	99	98
154	99	99	99	98
155	99	99	99	99
Based on sample size:	1008	2097*	1275	495

* This figure includes 312 from the IF comparability study (2011/12).

© 2003 – 2013 Team Focus Limited

NUMERICAL (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
60	1	1	1	1
61	2	1	1	1
62	2	1	1	1
63	2	1	1	1
64	2	1	1	1
65	2	1	1	1
66	3	1	1	1
67	4	1	1	1
68	5	1	1	1
69	5	1	1	1
70	6	1	1	1
71	7	1	1	1
72	8	1	1	1
73	9	1	1	1
74	10	1	1	1
75	11	1	1	1
76	12	1	1	1
77	13	1	1	1
78	15	3	1	1
79	17	4	1	1
80	19	5	1	1
81	21	6	2	1
82	24	7	2	1
83	27	8	3	1
84	30	10	3	1
85	34	13	4	1
86	37	15	5	1
87	39	16	5	1
88	43	18	6	1
89	46	20	7	1
90	49	23	8	1
91	53	25	10	1
92	56	28	11	1
93	59	33	12	3
94	62	37	13	4

© 2003 – 2013 Team Focus Limited

NUMERICAL (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
95	64	40	15	5
96	67	43	18	6
97	70	46	20	7
98	72	49	22	9
99	74	52	24	11
100	76	55	28	13
101	78	58	31	14
102	79	60	35	15
103	81	63	39	17
104	82	65	41	20
105	84	67	43	23
106	85	69	46	26
107	86	70	49	29
108	87	71	53	32
109	88	72	55	35
110	89	74	57	37
111	90	75	59	39
112	90	76	62	41
113	90	78	65	44
114	91	79	67	47
115	92	80	69	50
116	92	81	71	53
117	93	83	74	56
118	93	84	77	60
119	93	85	79	63
120	93	86	81	66
121	93	88	82	68
122	94	89	84	71
123	94	88	86	72
124	94	90	86	73
125	94	91	87	74
126	95	91	88	76
127	95	92	88	78
128	95	92	90	79
129	95	93	91	81
130	95	94	91	82
131	95	94	91	83

© 2003 – 2013 Team Focus Limited

NUMERICAL (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
132	96	94	92	84
133	96	95	92	85
134	96	95	92	86
135	96	96	94	87
136	96	96	94	88
137	97	96	95	89
138	97	96	96	90
139	97	96	96	91
140	97	97	96	91
141	97	97	96	91
142	97	97	97	92
143	97	97	97	92
144	97	98	97	93
145	97	98	98	94
146	98	99	98	94
147	98	99	98	95
148	98	99	98	96
149	98	99	98	96
150	98	99	99	96
151	98	99	99	96
152	98	99	99	97
153	99	99	99	97
154	99	99	99	97
155	99	99	99	97
156	99	99	99	98
Based on sample size:	**1773**	**3018***	**1817**	**930**

* This figure includes 358 from the IF comparability study (2011/12).

© 2003 – 2013 Team Focus Limited

© 2003 – 2013 Team Focus Limited

ABSTRACT (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
50	2	1	2	1
51	3	1	2	1
52	3	1	2	1
53	3	1	2	1
54	4	1	2	1
55	4	1	2	1
56	5	1	2	1
57	5	1	2	1
58	5	1	2	1
59	6	1	2	1
60	6	1	2	1
61	7	2	3	1
62	7	3	3	1
63	8	4	4	1
64	9	5	5	2
65	10	6	6	2
66	10	7	6	3
67	11	8	6	3
68	12	9	6	4
69	13	10	6	5
70	14	11	7	6
71	15	13	8	7
72	17	15	9	7
73	19	16	10	8
74	21	18	11	9
75	23	19	12	11
76	25	21	13	12
77	26	23	15	13
78	27	25	16	14
79	29	27	17	16
80	30	28	20	17
81	32	29	21	19
82	34	31	22	21
83	36	34	25	21
84	38	37	26	22
85	40	39	28	23

© 2003 – 2013 Team Focus Limited

ABSTRACT (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
86	42	41	31	24
87	44	43	32	25
88	47	46	34	26
89	51	48	35	28
90	54	51	37	29
91	57	53	40	30
92	58	55	43	32
93	59	57	44	33
94	61	59	46	35
95	62	61	48	37
96	64	64	50	39
97	66	66	52	42
98	68	69	55	43
99	69	70	57	47
100	70	72	59	51
101	72	74	62	53
102	74	76	64	55
103	79	77	66	56
104	81	79	70	58
105	83	80	71	62
106	84	81	72	63
107	85	82	73	65
108	86	83	75	66
109	87	84	77	68
110	88	85	78	70
111	89	86	80	71
112	89	87	80	73
113	90	88	81	75
114	91	90	82	77
115	92	90	84	78
116	93	91	85	80
117	94	92	86	81
118	94	93	88	83
119	95	94	88	84
120	96	94	89	85
121	96	95	90	86
122	96	95	90	88

© 2003 – 2013 Team Focus Limited

ABSTRACT (CLOSED) standardised IRT score	Percentiles			
	Level 1 GCSE	Level 2 A-Level	Level 3 UG	Level 4 PG
123	97	95	91	89
124	97	95	92	90
125	97	95	93	91
126	97	96	94	92
127	98	96	95	93
128	98	97	95	94
129	98	97	95	94
130	98	97	96	94
131	99	97	96	95
132	99	97	96	96
133	99	97	96	97
134	99	98	96	97
135	99	98	96	97
136	99	98	97	98
137	99	98	97	98
138	99	98	97	98
139	99	98	98	98
140	99	98	98	98
141	99	98	98	98
142	99	98	98	99
143	99	98	99	99
144	99	99	99	99
145	99	99	99	99
Based on sample size:	**768**	**1854***	**1226**	**469**

* This figure includes 329 from the IF comparability study (2011/12).

© 2003 – 2013 Team Focus Limited

Percentile to STEN conversion

If you want to estimate from a percentile to a STEN you can use the table below. For example, if a student has a percentile score of 30 it is equivalent to a STEN of 4; if it is 74 it is equivalent to a STEN of 7, and so on.

Percentile	STEN
0+	1
2+	2
7+	3
16+	4
31+	5
50+	6
69+	7
84+	8
93+	9
98+	10

© 2003 – 2013 Team Focus Limited